

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

LARISSA MEDEIROS DE ALMEIDA

AVALIAÇÃO DA GRAVIDADE DA MALÁRIA UTILIZANDO TÉCNICAS
DE EXTRAÇÃO DE CARACTERÍSTICAS E REDES NEURAIAS
ARTIFICIAIS

MANAUS

2015

UNIVERSIDADE FEDERAL DO AMAZONAS
FACULDADE DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

LARISSA MEDEIROS DE ALMEIDA

AVALIAÇÃO DA GRAVIDADE DA MALÁRIA UTILIZANDO TÉCNICAS
DE EXTRAÇÃO DE CARACTERÍSTICAS E REDES NEURAIIS
ARTIFICIAIS

Dissertação apresentada ao Curso de Mestrado em Engenharia Elétrica, área de concentração Controle e Automação de Sistemas do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas.

Orientadora: Prof^a. Dr^a. Marly Guimarães Fernandes Costa
Co-Orientador: Prof. Dr. Cícero Ferreira Fernandes Costa Filho

MANAUS

2015

LARISSA MEDEIROS DE ALMEIDA

AVALIAÇÃO DA GRAVIDADE DA MALÁRIA UTILIZANDO TÉCNICAS DE EXTRAÇÃO DE CARACTERÍSTICAS E REDES NEURAIAS ARTIFICIAIS.

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Amazonas, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica na área de concentração Controle e Automação de Sistemas.

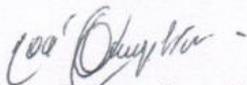
Aprovado em 17 de abril de 2015.

BANCA EXAMINADORA



Marly Guimarães Fernandes Costa, Presidente

Universidade Federal do Amazonas- UFAM



Prof. Dr. João Evangelista Neto, Membro

Universidade Federal do Amazonas- UFAM



Prof. Dr. Marco Antônio Gutierrez, Membro

Universidade de São Paulo-USP

M488a Medeiros de Almeida, Larissa
Avaliação da Gravidade da malária utilizando técnicas de extração de características e redes neurais artificiais / Larissa Medeiros de Almeida. 2015
127 f.: il. color; 29.7 cm.

Orientadora: Marly Guimarães Fernandes Costa
Coorientador: Cícero Ferreira Fernandes Costa Filho
Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Amazonas.

1. redes neurais. 2. regularização. 3. malária. 4. diagnóstico. 5. análise dos componentes principais. I. Costa, Marly Guimarães Fernandes II. Universidade Federal do Amazonas III. Título

AGRADECIMENTOS

Primeiramente, aos meus orientadores, Profa. Dra. Marly Guimarães Fernandes Costa e Prof. Dr. Cícero Ferreira Fernandes Costa Filho;

Aos meus pais e todos os meus familiares pelo apoio;

Aos meus colegas de turma Endrews, Jonilson, Márcio e Robson que sempre estiveram me incentivando e ajudando no desenvolvimento do trabalho;

Aos amigos Fábio Araújo, Fábio de Melo, Rafael Brown, Lucas Jardim e às amigas e ex-colegas da Samsung, Gabriele e Adriana que me incentivaram a largar tudo e seguir o meu sonho acadêmico;

Aos professores do curso de pós-graduação em engenharia elétrica pelo aprendizado adquirido;

À Universidade Federal do Amazonas e em especial ao Centro de Tecnologia Eletrônica e da Informação – CETELI - pela concessão de toda infraestrutura para a realização desse trabalho;

A Capes pela concessão da bolsa de estudos no primeiro ano de trabalho;

À Samsung. Parte dos resultados apresentados neste trabalho foram obtidos através do Projeto de pesquisa e formação de recursos humanos, em nível de graduação e pós-graduação, nas áreas de automação industrial, softwares para dispositivos móveis e TV Digital, financiado pela Samsung Eletrônica da Amazônia Ltda., no âmbito da Lei no. 8.387 (art. 2º) /91.

RESUMO

Cerca de metade da população mundial vive em áreas de risco da malária. Além disso, dada a globalização das viagens, essas doenças que antes eram consideradas exóticas e principalmente tropicais são cada vez mais encontradas em salas de emergência de hospitais no mundo todo. E frequentemente quando se trata de experiência em doenças tropicais, a opinião de especialistas na maioria das vezes está indisponível ou não acessível em tempo hábil. A tarefa de chegar a um diagnóstico da malária preciso e eficaz, fundamental na prática médica, pode tornar-se complexa. E a complexidade desse processo aumenta à medida que os pacientes apresentam sintomas não específicos com uma grande quantidade de dados e informação imprecisa envolvida. Nesse sentido, Uzoka e colaboradores (2011a), a partir de informações clínicas de 30 pacientes nigerianos com diagnóstico confirmado de malária, utilizaram a metodologia *Analytic Hierarchy Process* (AHP) e metodologia *Fuzzy* para realizar a avaliação da gravidade da malária. Os resultados obtidos foram comparados com o diagnóstico de médicos especialistas. Esta dissertação desenvolve uma nova metodologia para avaliação da gravidade da malária e a compara com as técnicas utilizadas por Uzoka e colaboradores (2011a). Para tal o conjunto de dados utilizados é o mesmo do referido estudo. A técnica utilizada é a de Redes Neurais Artificiais (RNA). São avaliadas três arquiteturas com diferentes números de neurônios na camada escondida, duas metodologias de treinamento (*leave-one-out* e *10-fold cross-validation*) e três critérios de parada, a saber: o erro médio quadrático, parada antecipada e regularização. Na primeira fase, é utilizado o banco de dados completo. Posteriormente, são utilizados os métodos de extração de características: na segunda fase, a Análise dos Componentes Principais (do inglês, *Principal Component Analysis* - PCA) e na terceira fase, a Análise Discriminante Linear (do inglês, *Linear Discriminant Analysis* - LDA). O melhor resultado obtido nas três fases, foi com o banco de dados completo, utilizando o critério de regularização, associado ao *leave-one-out*, de 83.3%. Já o melhor resultado obtido em (Uzoka, Osuji e Obot, 2011) foi com a rede *fuzzy* onde obteve 80% de acurácia.

Palavras-chave: Redes Neurais Artificiais; Regularização; Parada Antecipada; Análise dos Componentes Principais; Análise Discriminante Linear; malária; diagnóstico.

ABSTRACT

About half the world's population lives in malaria risk areas. Moreover, given the globalization of travel, these diseases that were once considered exotic and mostly tropical are increasingly found in hospital emergency rooms around the world. And often when it comes to experience in tropical diseases, expert opinion most of the time is not available or not accessible in a timely manner. The task of an accurate and efficient diagnosis of malaria, essential in medical practice, can become complex. And the complexity of this process increases as patients have non-specific symptoms with a large amount of data and inaccurate information involved. In this approach, Uzoka and colleagues (2011a), from clinical information of 30 Nigerian patients with confirmed malaria, used the Analytic Hierarchy Process method (AHP) and Fuzzy methodology to conduct the evaluation of the severity of malaria. The results obtained were compared with the diagnosis of medical experts. This paper develops a new methodology to evaluate the severity of malaria and compare with the techniques used by Uzoka and colleagues (2011a). For this purpose the data set used is the same of that study. The technique used is the Artificial Neural Networks (ANN). Are evaluated three architectures with different numbers of neurons in the hidden layer, two training methodologies (*leave-one-out* and *10-fold cross-validation*) and three stopping criteria, namely: the root mean square error, early stop and regularization. In the first phase, we use the full database. Subsequently, the feature extraction methods are used: in the second stage, the Principal Component Analysis (PCA) and in the third stage, the Linear Discriminant Analysis (LDA). The best result obtained in the three phases, it was with the full database, using the criterion of regularization associated with the leave-one-out method, of 83.3%. And the best result obtained in (Uzoka, Osuji and Obot, 2011) was with the fuzzy network which revealed 80% accuracy.

Keywords: Artificial Neural Networks; Regularization; Early stop; Principal Component Analysis; Linear discriminant analysis; malaria; diagnosis.

LISTA DE ILUSTRAÇÕES

Figura 1 – População mundial com risco de malária. As áreas em cinza claro não possuem um risco relevante.....	12
Figura 2 – Módulos Principais do IMEX. Fonte: Adaptado de (Suan, 1990).....	21
Figura 3 – (a) Módulo de Inicialização e (b) Módulo de Diagnóstico. Fonte: Adaptado de (Suan, 1990).....	22
Figura 4 - Visualização da rede de surtos de doenças diagnosticadas com o potencial de causar encefalite (colorido) e surtos de encefalite não diagnosticada (branco). Fonte: Adaptado de (Bogich <i>et al.</i> , 2013).....	29
Figura 5- PWC do nível 2 (Uzoka, Osuji e Obot, 2011).	34
Figura 6 - PWC do nível 3 (Uzoka, Osuji e Obot, 2011).	35
Figura 7 - DFI1 Prioridades do nível 2(Uzoka, Osuji e Obot, 2011)	37
Figura 8 - DFI2 Prioridades do nível 3(Uzoka, Osuji e Obot, 2011)	37
Figura 9 - Esquema ilustrativo de um sistema fuzzy. Fonte: Adaptado de (Schiassi <i>et al.</i> , 2008; Ferreira, 2009)	41
Figura 10 - Modelo não linear de um neurônio. Fonte: Adaptado de (Haykin e Network, 2004)	46
Figura 11 - Fluxo de processamento do algoritmo backpropagation. Fonte: Adaptado de (Caudill e Butler, 1992)	47
Figura 12 - Função seno com ruído representada pela linha tracejada e uma função produzida por uma rede neural 1-20-1 representada pela linha contínua. Fonte: Adaptado de The MathWorks, Inc. (2008)	51
Figura 13 - Curva do Erro de Treinamento e Erro de Validação. Eixo vertical: Erro. Eixo Horizontal: Tempo. Fonte: Adaptado de (Prechelt, 1998)	52
Figura 14 - Função triangular com 100 pontos de ruído (representados por +) e uma função produzida por rede neural representada pela linha curva. No lado esquerdo, a rede neural sem regularização e no lado direito, a rede neural com regularização bayesiana. Fonte: Adaptado de Foresee e Hagan (1997).	59
Figura 15- Três fases da pesquisa e os diferentes conjuntos de dados utilizados	68
Figura 16 – Diagrama em blocos genérico aplicado em cada fase da pesquisa. cd_i representa o número de neurônios na entrada (igual ao número de características, podendo valer 10, 15 ou 22) e n_{hi} representa o número de neurônios na camada intermediária.	69
Figura 17 - Representação dos critérios de parada extraídos da Figura 16	72
Figura 18 - Metodologias de treinamento e teste extraído da Figura 16	75
Figura 19 - <i>10-fold Cross Validation</i>	77
Figura 20 - <i>Leave-one-out</i>	78
Figura 21 – As quatro arquiteturas utilizadas para o treinamento.....	80

LISTA DE TABELAS

Tabela 1 – Revisão de diversos estudos que exploram preditores de malária clínica em crianças realizada por (Chandramohan, Jaffar e Greenwood, 2002).....	24
Tabela 2 - Sensibilidade e especificidade de preditores potenciais da malária em crianças. Fonte: (Chandramohan, Jaffar e Greenwood, 2002).....	25
Tabela 3 - Resultados obtidos por (Barros <i>et al.</i> , 2010).....	27
Tabela 4 – Hierarquia dos critérios básicos do diagnóstico da malária. Fonte: Adaptado de (Uzoka, Osuji e Obot, 2011).....	33
Tabela 5 - Escala de Intensidade de Malária (Uzoka, Osuji e Obot, 2011).....	38
Tabela 6 - Sumário dos resultados apresentados por Uzoka, Osuji e Obot . Fonte: Adaptado de (Uzoka, Osuji e Obot, 2011).....	41
Tabela 7 - Análise Comparativa das implementações AHP e <i>Fuzzy</i> dos valores dos diagnósticos normalizados.....	42
Tabela 8 – Principais artigos abordados na revisão bibliográfica	44
Tabela 9 – Sumário dos passos metodológicos a serem realizados para cada uma das 4 arquiteturas de redes neurais utilizadas	70
Tabela 10 - Detalhamento da divisão dos grupos nos 10 experimentos segundo a técnica <i>10-fold-cross-validation</i>	76
Tabela 11 - Acurácia dos métodos propostos na primeira fase	89
Tabela 12 – Resumo dos resultados obtidos na segunda fase com 15 características.....	95
Tabela 13 - Resumo dos resultados obtidos na segunda fase com 10 características	101
Tabela 14 - Resumo dos resultados obtidos na terceira fase utilizando LDA com 15 características.....	106
Tabela 15 - Resumo dos resultados obtidos na terceira fase utilizando LDA com 10 características.....	110
Tabela 16 - Melhores resultados de cada critério de parada.....	112
Tabela 17 - Comparação dos métodos de avaliação da gravidade da malária, incluindo diagnóstico médico (Uzoka, Osuji e Obot, 2011), redes neurais regularizadas e o método <i>fuzzy</i> (Uzoka, Osuji e Obot, 2011).....	113

SUMÁRIO

1	INTRODUÇÃO.....	11
1.1	OBJETIVOS	15
1.1.1	GERAL.....	15
1.1.2	ESPECÍFICOS.....	15
2	REVISÃO BIBLIOGRÁFICA	18
3	FUNDAMENTAÇÃO TEÓRICA	45
3.1	REDES NEURAIS ARTIFICIAIS	45
3.1.1	Aprendizado Supervisionado: O Algoritmo Backpropagation.....	47
3.2	OVERFITTING.....	50
3.3	TÉCNICAS DE REDUÇÃO DE DIMENSIONALIDADE	60
4	MATERIAIS E MÉTODOS.....	67
4.1	MATERIAIS	67
4.2	MÉTODOS.....	68
5	RESULTADOS	82
	REFERÊNCIAS	115

INTRODUÇÃO

A malária é uma doença grave, potencialmente causada por parasitas *Plasmodium*. Existem mais de cem tipos de plasmódio. Dos que infectam o homem, quatro são os mais importantes: *Plasmodium vivax*, *Plasmodium falciparum*, *Plasmodium malariae* e *Plasmodium ovale*. Plasmódios são transmitidos entre seres humanos pela picada do mosquito *Anopheles* que carrega o parasita. O ciclo da malária humana é homem-anofelino-homem. Geralmente é a fêmea que ataca o homem porque precisa de sangue para garantir a desova e o amadurecimento dos ovos. Depois de picar um indivíduo infectado, o parasita desenvolve parte de seu ciclo no mosquito e, quando alcança as glândulas salivares do inseto, está pronto para ser transmitido para outra pessoa (Camargo, 2003).

Quando os plasmódios humanos entram na corrente sanguínea, viajam até o fígado e se reproduzem rapidamente. Na maioria das formas de malária, alguns parasitas vão para a corrente sanguínea e também destroem as células vermelhas do sangue, que transportam oxigênio para os tecidos vitais do corpo, podendo resultar em anemia (Ferreira, 2013).

Os parasitas que ficam no fígado, continuam a se reproduzir e a enviar mais parasitas para a corrente sanguínea. Isso resulta em ataques repetidos de malária, que se inicia com sintomas de gripe, e cada vez mais novos parasitas são liberados para o sangue. Esses ataques podem ocorrer por anos se a doença não for diagnosticada e tratada. Eventualmente, o sistema imunológico do corpo pode desenvolver uma defesa contra tais ataques, os quais podem se tornar menos severos em algumas pessoas (Ministério da Saúde, 2005).

A malária nos países tropicais e subtropicais é uma séria ameaça à saúde pública. Cerca de metade da população mundial vive em áreas de risco, como pode se observar na Figura 1. No

Brasil, de 2000 a 2011, segundo dados do Ministério da Saúde, 99,7% do total de casos de malária são registrados na Amazônia Legal, composta pelos estados do Acre, Amazonas, Amapá, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins.

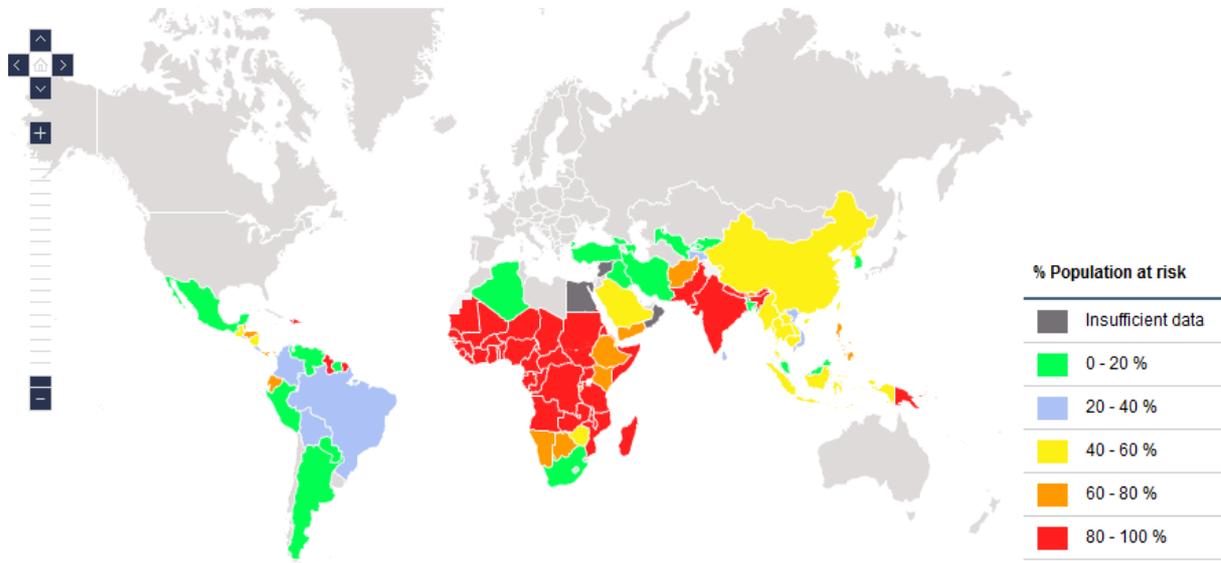


Figura 1 – População mundial com risco de malária. As áreas em cinza claro não possuem um risco relevante.
Fonte: Global Malaria Mapper – WHO World Malaria Report, World Health Organization (2012).

Dependendo do tipo de plasmódio, o momento do surgimento dos sintomas após a picada do mosquito pode variar. Com o *Plasmodium falciparum*, os sintomas no hospedeiro demoram de 7 a 14 dias para aparecer. Enquanto que com o *Plasmodium vivax* e *ovale*, a demora é de 8 a 14 dias e o *Plasmodium malariae*, de 7 a 30 dias. Há também casos relatados em que os sintomas da infecção da malária são desenvolvidos 12 meses após o paciente ter sido picado por um mosquito, pois o plasmódio pode permanecer dormente no fígado por um longo período (Ministério da Saúde, 2005). Os sintomas são semelhantes aos da gripe e dependendo do estágio da doença tipicamente incluem: calafrios, fadiga, febre, dor de cabeça, náuseas, vômitos, falta de apetite, sudorese, dor muscular nas pernas e costas, dor abdominal, diarreia, tonturas, tosse, pele amarelada e cansaço (Camargo, 2003). Os sintomas da infecção por malária nem sempre são dramáticos. Dessa forma uma pessoa com os referidos sintomas, pode algumas vezes ser diagnosticada com algo sem importância. No entanto, se a doença não for tratada, a deterioração

normalmente é súbita e drástica com os parasitas aumentando rapidamente dentro da corrente sanguínea.

O sucesso terapêutico da malária está ligado a um diagnóstico adequado. A Organização Mundial de Saúde (*World Malaria Report*, 2012) recomenda que a busca de diagnósticos precoces e precisos deve ser uma ação prioritária que potencializa a redução das complicações e minimização de custos.

Ferreira (2013), identifica as seguintes complicações da malária: a malária cerebral, a transmissão materno-infantil, baixo peso ao nascer, anemia, icterícia, insuficiência renal, desequilíbrio de fluidos, aumento do baço, aumento do fígado, complicações hepáticas, complicações do cérebro e a morte. Outras complicações incluem (Mehta *et al.*, 2001): coagulação intravascular disseminada (DIC) e síndrome do desconforto respiratório agudo (ARDS). Em muitos casos, a distinção entre os sintomas de malária e de complicações da malária não é clara ou é arbitrária.

Tradicionalmente, o diagnóstico da malária é realizado através do exame laboratorial da gota espessa, recomendado pela OMS e também pelo Ministério da Saúde (*World Malaria Report*, 2012), o qual consiste na microscopia da gota espessa de sangue, colhida por punção digital e corada pelo método de Walker, através da contagem do número de parasitas, em campos de grande aumento, por 100 leucócitos, para em seguida se determinar a densidade da parasitemia em relação aos campos microscópicos examinados. Outra forma de diagnóstico da malária além da microscopia de gota espessa é a molecular. O diagnóstico molecular, por sua vez, só começou a ser testado em meados dos anos 80. A técnica molecular mais utilizada é o Nested PCR (*Polymerase-Chain Reaction* ou reação em cadeia da polimerase). O PCR é uma técnica que se baseia na multiplicação da sequência genética do parasita e que pode produzir resultados excelentes com elevada especificidade e sensibilidade. No entanto, em virtude do alto custo do aparato tecnológico envolvido e da necessidade de infraestrutura e mão de obra

especializada seu uso ainda é restrito a laboratórios de referência e, portanto, difícil de ser utilizado pelas próprias características das mesmas (Barros *et al.*, 2010).

Doenças tropicais como a malária são na maioria das vezes consideradas condições regionais, pois a disponibilidade de recursos dedicados ao seu diagnóstico e tratamento é escassa no mundo desenvolvido, onde esta condição não é endêmica. No entanto, dada a globalização das viagens, essas doenças que antes eram consideradas exóticas e principalmente tropicais são cada vez mais encontradas em salas de emergência de hospitais no mundo todo (Camargo, 2003). Consultas médicas especializadas têm sido usadas para preencher esta lacuna, mas frequentemente quando se trata de experiência em doenças tropicais, a opinião de especialistas na maioria das vezes está indisponível ou não acessível em tempo hábil.

A tarefa de chegar a um diagnóstico da malária preciso e eficaz, fundamental na prática médica, pode tornar-se, por vezes, complexa e de difícil controle. A complexidade desse processo aumenta a medida que os pacientes apresentam sintomas não específicos com uma grande quantidade de dados e informação imprecisa envolvida (Seising, 2006). Bronzan, Mcmorrow e Kachur (2008) afirmam que erros diagnósticos levam a uma gestão pobre da doença e ao aumento da morbidade e mortalidade. Esta situação é agravada pela elevada taxa de pobreza, falta de infraestrutura educacional e hospitalar, manutenção de registros médicos pobres, má formação e falta de um número suficiente de médicos.

O domínio de interesse neste trabalho é a avaliação da gravidade em pacientes com sintomas clínicos ou manifestações de malária em um cenário de sintomas ambíguos e comuns a várias outras patologias. Nesse sentido, Uzoka e colaboradores (2011a), a partir de informações clínicas de 30 pacientes nigerianos com diagnóstico confirmado de malária, utilizou a metodologia *Analytic Hierarchy Process* (AHP) e metodologia *fuzzy* para realizar o estadiamento da malária em cada um dos 30 pacientes. Os resultados obtidos foram comparados com o diagnóstico de médicos especialistas.

O trabalho proposto nessa dissertação tem por escopo o desenvolvimento de uma nova metodologia para **avaliação da gravidade da malária** e a comparação desta com as técnicas utilizadas por Uzoka e colaboradores (2011a). Para tal o conjunto de dados utilizados é o mesmo do referido estudo. Não é do escopo deste trabalho a realização do diagnóstico diferencial dos pacientes, em que se identifica a presença ou ausência de determinada doença.

A técnica utilizada é a de Redes Neurais Artificiais (RNA). Serão avaliadas três arquiteturas e três critérios de parada, a saber: o erro médio quadrático, parada antecipada e regularização.

1.1 OBJETIVOS

1.1.1 GERAL

Avaliar e comparar diferentes metodologias utilizando redes neurais para auxílio a avaliação da gravidade da malária em pacientes com diagnóstico confirmado.

1.1.2 ESPECÍFICOS

- Identificar as principais ferramentas de Suporte à Decisão Clínica em Malária publicadas, através de uma pesquisa bibliográfica.
- Caracterizar a utilização de redes neurais, com critério de parada utilizando o erro médio quadrático, na avaliação da gravidade da malária.
- Caracterizar a utilização de redes neurais com aprendizado supervisionado do tipo *backpropagation*, com critério de parada antecipado, na avaliação da gravidade da malária.

- Caracterizar a utilização de redes neurais com aprendizado supervisionado do tipo *backpropagation*, com critério de parada utilizando a regularização, na avaliação da gravidade da malária.
- Utilizar os métodos de extração de características: Análise dos Componentes Principais (do inglês, *Principal Component Analysis - PCA*) e Análise Discriminante Linear (do inglês, *Linear Discriminant Analysis – LDA*) associados a redes neurais na avaliação da gravidade da malária.

1.2 ORGANIZAÇÃO

Essa dissertação de mestrado está dividida nos seguintes capítulos:

- Introdução (Capítulo 1);
- Revisão Bibliográfica (Capítulo 2);
- Fundamentação Teórica (Capítulo 3);
- Materiais e Métodos (Capítulo 4);
- Resultados (Capítulo 5);
- Discussões e Conclusões (Capítulo 6);
- Referências Bibliográficas;
- Apêndice.

O Capítulo 2 consistirá em identificar os sistemas de suporte à decisão clínica em malária até então publicados, através de uma pesquisa bibliográfica nas principais bases de dados literárias. No Capítulo 3, apresentamos a fundamentação teórica necessária ao desenvolvimento do presente trabalho através da descrição das ferramentas utilizadas.

No Capítulo 4, todo o conjunto de dados e os passos metodológicos utilizados são apresentados. No Capítulo 5 são exibidos os resultados obtidos com o método proposto, tais

como: desempenho dos classificadores neurais e taxas de acertos do método. Apresentam-se, também, tabelas de confusão e gráficos comparativos. Por fim, conclusões do trabalho são apresentadas e sugerem-se trabalhos futuros.

No Apêndice, encontra-se cópia do artigo publicado intitulado: “AVALIAÇÃO DA GRAVIDADE DA MALÁRIA UTILIZANDO REDES NEURAIIS: UM ESTUDO COMPARATIVO” de Larissa Almeida, Cícero Costa Filho e Marly Costa Filho. Apresentado no XXIV Congresso Brasileiro de Engenharia Biomédica, realizado na cidade de Uberlândia, Minas Gerais, de 13 a 17 de outubro de 2014.

REVISÃO BIBLIOGRÁFICA

Como ferramenta potencial de tratamento das ambiguidades e incertezas associadas com os sintomas e dados clínicos, surgem os sistemas de apoio à decisão clínica (CDSS, do inglês *Clinical Decision Support System*), conhecidos por melhorar o desempenho profissional, reduzir custos e apoiar diagnósticos mais precisos (Seising, 2006).

Vários são os sistemas de apoio a decisão clínica publicados na literatura, dentre os quais se citam (Ilfie *et al.*, 2002; Ramnarayan e Britto, 2002; Weir *et al.*, 2003; Toth-Pal *et al.*, 2008; Bogich *et al.*, 2013). Poucos são dedicados ao diagnóstico da malária (Suan, 1990; Barros *et al.*, 2010; Uzoka, Osuji e Obot, 2011). Um número considerável desses tentou mitigar os desafios associados ao desenvolvimento de CDSS se debruçando nas tarefas de aquisição de conhecimento, representação e utilização no diagnóstico médico. No entanto, o problema da gestão do conhecimento impreciso ainda perdura. Neste capítulo abordaremos alguns desses sistemas, identificados na literatura.

2.1 PESQUISA BIBLIOGRÁFICA

Foi realizada uma pesquisa bibliográfica nas seguintes bases de dados literárias: IEEEExplorer, *Engineering Village* e PubMed. Após a leitura inicial, foram selecionados os artigos mais relevantes e fez-se outra busca através dos vocabulários indexados a eles. A partir da análise desses artigos, foram identificados diversos Sistemas de Apoio ao Diagnóstico, com as mais variadas técnicas e cada um com características bem peculiares, os quais serão apresentados a seguir.

2.2 SISTEMAS DE APOIO À DECISÃO CLÍNICA (CDSS's)

Segundo (Garg *et al.*, 2005), sistemas de apoio à decisão clínica (CDSSs) são sistemas de informação destinados a melhorar a tomada de decisão clínica. As características dos pacientes são comparadas com uma base de conhecimento informatizada e algoritmos de *software* geram recomendações específicas sobre o paciente.

Esses sistemas fornecem vários modos de apoio à decisão, incluindo conselhos para a prescrição de medicamentos, alertas de valores críticos e de saúde preventiva, além de registros médicos com o histórico do paciente. É um campo que está evoluindo rapidamente por causa dos avanços tecnológicos, da facilidade de acesso a sistemas computadorizados na prática clínica, e da crescente preocupação sobre o processo e qualidade dos cuidados médicos. Mas, como acontece com qualquer inovação nos cuidados de saúde, CDSSs devem ser rigorosamente avaliados e padronizados antes de sua utilização na prática clínica (Garg *et al.*, 2005).

Em (Ahmadian *et al.*, 2011) foi realizada uma revisão da literatura para investigar se Sistemas de Suporte a Decisão Clínica (CDSS) usam dados padronizados (codificados ou numéricos) e quais sistemas terminológicos têm sido usados para a codificação. Os artigos foram identificados com base em uma revisão da literatura sobre o CDSS anterior e em estudos identificados pela *American Medical Informatics Association (AMIA)*'s "Year in Review" de outubro de 2004 até outubro de 2008, totalizando 77 artigos. Cerca de 22% dos CDSSs avaliados usam somente dados numéricos e 51% dos CDSSs utilizam dados codificados numa terminologia internacional. A terminologia internacional mais utilizada (68% dos casos) foi o ICD (do inglês) ou CID (Classificação Internacional de Doenças e Problemas Relacionados à Saúde), seguida pelo LOINC (*Logical Observation Identifiers Names and Codes*), em 12% dos casos.

A falta de dados padronizados é mencionada pela maioria dos pesquisadores como um grande obstáculo para o desenvolvimento, implementação e compartilhamento dos CDSSs. Esta diversidade dificulta a possibilidade de partilha e de raciocínio com dados dentro de diferentes sistemas. Essa pesquisa confirma a hipótese de que a padronização de dados é um fator crítico de sucesso para o desenvolvimento CDSS.

2.3 SISTEMAS DE APOIO AO DIAGNÓSTICO DA MALÁRIA

Datam de mais de duas décadas as primeiras publicações científicas que abordam o desenvolvimento de ferramentas de suporte ao diagnóstico da malária. Uma das primeiras propostas de sistemas de auxílio ao diagnóstico e tratamento da malária, sistematizando o processo de tomada de decisão médica em um programa de computador foi apresentada por (Suan, 1990), o sistema IMEX (*Integrated Malaria Express*). Dispõe de uma interface amigável com o usuário, o sistema inicialmente coleta informações clínicas do paciente que permitem a um sistema baseado em regras inferir se o paciente encontra-se ou não infectado com o *plasmodium*. Em caso positivo, o módulo de avaliação da gravidade da malária é acionado. Para esse módulo, novas informações são coletadas, como o resultado da contagem do número de parasitas. Dependendo da gravidade da malária, o sistema emite então recomendações adequadas para o tratamento.

Acoplado ao módulo de tratamento, existe o módulo de informação sobre as drogas. Este oferece informações atualizadas sobre os medicamentos antimaláricos comuns e é uma fonte conveniente de informações para o usuário.

Em resumo, o sistema IMEX é implementado em quatro módulos. Cada módulo tem a sua própria base de conhecimento que é organizado como uma hierarquia de regras, e passa conclusões intermediárias para os outros módulos. O módulo de **inicialização** executa as

funções preliminares como: inicializar as informações pessoais do paciente (para novos pacientes), ou obter informações a partir do banco de dados do paciente (para um paciente já cadastrado) e exibir para o usuário as informações de diagnóstico e tratamentos prévios. Normalmente, os módulos de **diagnóstico** e de **tratamento** são chamados em sequência, mas, se o diagnóstico já foi feito (manualmente ou através do IMEX), então o referido módulo é ignorado. A partir daí, o sistema pede detalhes da situação do paciente e utiliza esses dados no módulo de tratamento subsequente.

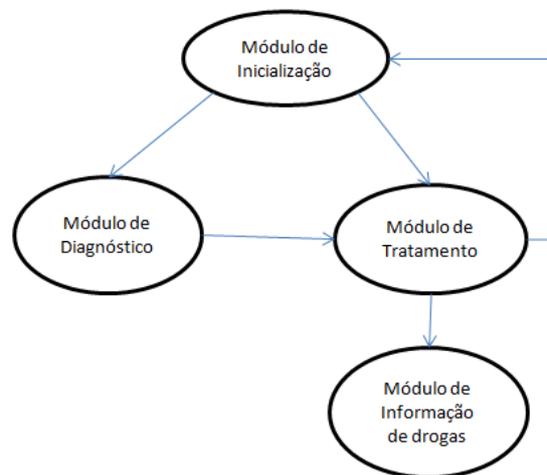


Figura 2 – Módulos Principais do IMEX. Fonte: Adaptado de (Suan, 1990)

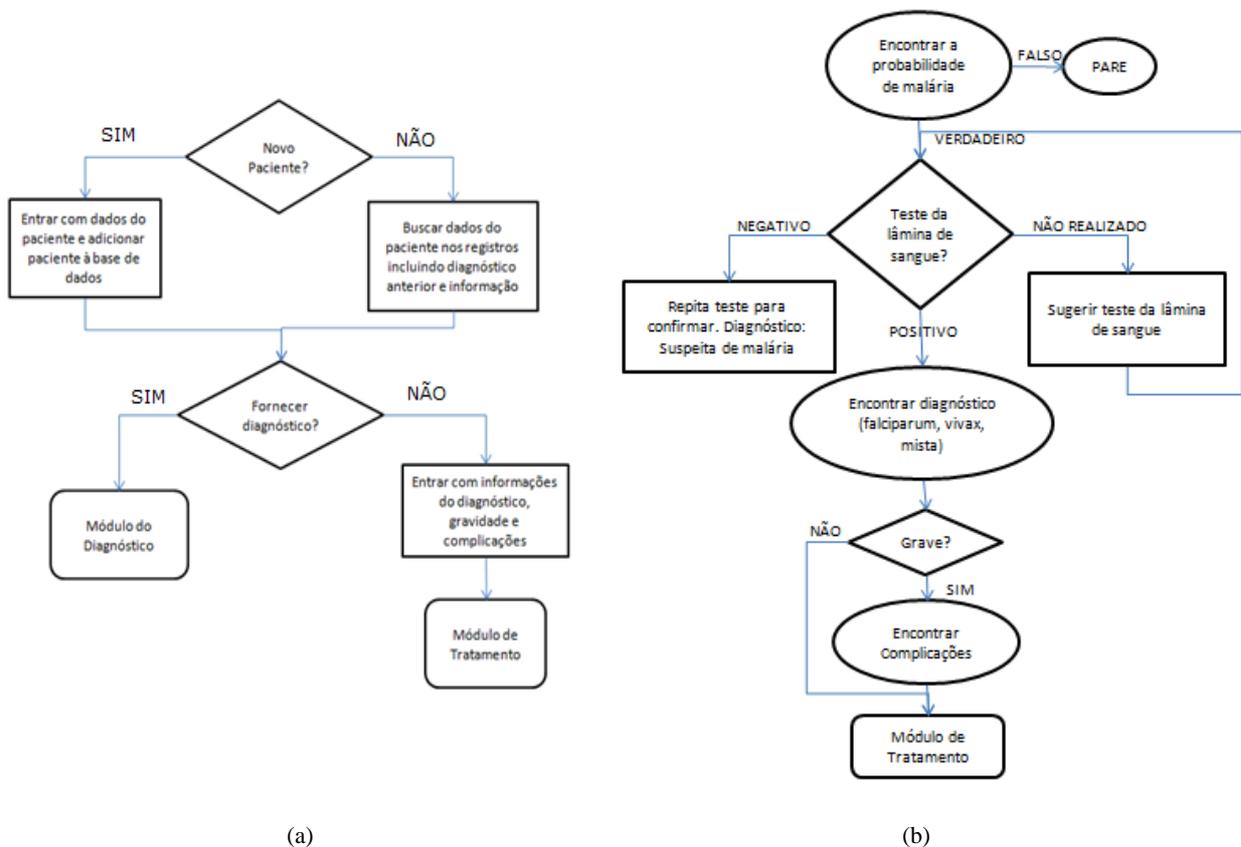


Figura 3 – (a) Módulo de Inicialização e (b) Módulo de Diagnóstico. Fonte: Adaptado de (Suan, 1990)

O sistema IMEX foi desenvolvido em uma época em que alguns dos usuários finais (médicos) nunca tinham usado um computador, mas foi considerado relativamente fácil de operar por eles. O módulo de diagnóstico para a detecção da malária é suficientemente abrangente, embora o processo para o diagnóstico de várias complicações é demasiado genérico. Portanto, um caso muito grave e complicado justificaria o auxílio de um especialista humano. No entanto, o *design* modular do IMEX permitiria que novos conhecimentos fossem facilmente adicionados ou modificados. E conseqüentemente, ele “auto-especializaria” com o passar do tempo.

2.3.1 Metodologias empregadas nos CDSS's aplicados a malária

O intuito principal do CDSS é estabelecer paralelismo entre os pacientes novos e um banco de dados prévio, de modo que, apoiados em diagnósticos anteriores, faz-se um diagnóstico aproximado sobre o agente causador da doença. Assim, quanto mais dados são solicitados, mais precisão será. O que difere esses sistemas são as metodologias empregadas no desenvolvimento da inteligência computacional.

Posteriormente a realização do estudo de Suan (1990), surgiram outros esforços no desenvolvimento desse tipo de aplicação, sempre com a finalidade de melhorar a qualidade dos cuidados médicos em pacientes com malária.

Chandramohan *et al.* (2001) apresentam um algoritmo que visa identificar os melhores preditores da malária através de métodos estatísticos. Através desse preditores, encontra-se o diagnóstico mais provável da malária. Este sistema foi testado com informações diagnósticas de 1.945 crianças e 2.885 adultos que apresentavam febre. Esses dados foram coletados entre junho de 1997 e janeiro de 1999 de moradores de uma área urbana da Índia. Todos os pacientes que participaram da clínica foram submetidos ao exame laboratorial de microscopia da gota espessa de sangue e exame físico de rotina. Foram registradas listas com sintomas e sinais dos pacientes. O padrão-ouro adotado foi a densidade da parasitemia da lâmina.

Para cada paciente avaliado foi calculado um grau para gravidade da malária, através da simples contagem dos sintomas e sinais presentes em uma lista previamente elaborada, com base na análise de significância estatística de avaliações clínicas anteriores. Das 1.945 crianças incluídas na análise, 139 (7%) tiveram malária (22 *Plasmodium falciparum*, 115 *Plasmodium vivax* e duas infecções mistas) e dos 2.885 adultos, 349 (12%) tiveram malária (70 *Plasmodium falciparum*, 266 *Plasmodium vivax* e 13 infecções mistas). As distribuições de idade por categoria de diagnóstico foram semelhantes: crianças com malária tinham uma idade média de 6

anos (variação de 1 a 14) em comparação com 5 anos para os casos de não-malária e ambos os conjuntos de adultos tinham uma idade média de 28 anos.

Os sintomas que se apresentam comuns, como febre intermitente, calafrios, sudorese, tremores, dores nas articulações ou dores de cabeça não tiveram individualmente um nível aceitável de sensibilidade e especificidade para o diagnóstico da malária. A ausência de tosse teve o melhor poder discriminante (sensibilidade 64%, especificidade de 56%) em crianças, e a presença de calafrios (sensibilidade 71% e especificidade de 40%) o melhor poder discriminante em adultos.

No ano seguinte, Chandramohan e colaboradores em (Chandramohan, Jaffar e Greenwood, 2002) revisaram os resultados de estudos publicados até então (Rooth e Björkman, 1992; Genton *et al.*, 1994; Gomes *et al.*, 1994; Redd *et al.*, 1996; Weber *et al.*, 1997; Luxemburger *et al.*, 1998; Olaleye *et al.*, 1998; Muhe *et al.*, 1999; Chandramohan *et al.*, 2001) a fim de avaliar os riscos e benefícios desses algoritmos baseados na identificação dos preditores (sintomas e sinais) com métodos estatísticos para o diagnóstico da malária em diferentes contextos epidemiológicos.

A Tabela 1 apresenta um resumo de alguns deles, incluindo local de estudo, critérios adotados, padrão ouro e tamanho da população.

Tabela 1 – Revisão de diversos estudos que exploram preditores de malária clínica em crianças realizada por (Chandramohan, Jaffar e Greenwood, 2002)

Referência	Local	Critérios	Padrão ouro	Pacientes	
				Com malária	Sem malária
Rooth e Bjorkman (1992)	Tanzânia	Crianças com histórico de febre	Parasitemia e temperatura \geq 37.5°C	395	273
Gomes et al. (1994)	Filipinas	Crianças febris ou com histórico de febre	Parasitemia e febre ou histórico de febre	97	47
Genton et al. (1994)	Papua Nova Guiné	Crianças com suspeita de malária	Parasitemia e febre ou histórico de febre	1077	406
Redd et al. (1996)	Malauí	Crianças com qualquer doença	Parasitemia e febre ou histórico de febre	672	452

Weber et al. (1997)	Gâmbia	Crianças com qualquer doença	Parasitemia e febre ou histórico de febre	28	379
Luxemburger et al (1998)	Tailândia	Crianças com febre ou histórico de febre nos últimos 3 dias	Parasitemia e febre ou histórico de febre	299	299
Muhe et al. (1999)	Etiópia	Crianças febris	Parasitemia e febre	48	756
Olaleye et al. (1998)	Gâmbia	Crianças febris nos últimos 3 dias com temperatura \geq 37.5°C	Parasitemia acima de 5000 microlitros e temperatura \geq 37.5°C	189	248

Em muitas áreas endêmicas de malária, a maioria dos pacientes com doença febril aguda recebem medicamentos antimaláricos de cuidadores e farmacêuticos. Neste caso, a sensibilidade do diagnóstico de malária estaria perto de 100%, mas a especificidade do seu diagnóstico é provável que seja muito baixo. Prescrever tratamento para a malária com um baixo limiar de suspeita é bom quando os parasitas da malária são sensíveis à cloroquina, já que a cloroquina é segura e barata. No entanto, é necessário um elevado grau de precisão no diagnóstico de zonas de resistência à cloroquina, devido à necessidade de tratamento com medicamentos mais caros e potencialmente mais perigosos. Na maioria desses estudos revisados, o padrão ouro foi o exame laboratorial de microscopia da gota espessa de sangue.

Outra observação interessante extraída desta pesquisa é que os melhores valores preditivos para o diagnóstico de malária variaram entre as populações, como pode ser observado na Tabela 2. Isto pode ter sido devido a diferenças nas percepções culturais, a endemicidade da malária, espécies de malária ou experiência clínica de investigadores entre os locais de estudo.

Tabela 2 - Sensibilidade e especificidade de preditores potenciais da malária em crianças. Fonte: (Chandramohan, Jaffar e Greenwood, 2002)

Preditores	Gâmbia [1]		Gâmbia [2]		Índia [3]		Tailândia [4]	
	Sen. (%)	Esp. (%)	Sen. (%)	Esp. (%)	Sen. (%)	Esp. (%)	Sen. (%)	Esp. (%)
Fome reduzida	72	49	90	23	27	76	NR	NR
Sonolência	82	34	67	51	1	99	NR	NR

Calafrios / arrepios	72	47	68	51	49	57	NR	NR
Ausência de tosse	65	60	37	73	54	64	60	53
Febre intermitente	NR	NR	NR	NR	96	9	NR	NR
Vômito	NR	NR	69	51	24	81	29	87
Dor de cabeça	NR	NR	NR	NR	37	68	83	42
Temperatura $\geq 38^{\circ}\text{C}$	NR	NR	NR	NR	41	71	51	71
Palidez na conjuntiva	33	84	33	71	32	68	NR	NR

NR = Não Reportado; Sen. = sensibilidade; Esp. = especificidade; [1] Olaleye *et al.* (1998); [2] Bojang *et al.* (2000); [3] Chandramohan *et al.* (2001); [4] Luxumburger *et al.* (1998)

Em pesquisas mais recentes, outros tipos de metodologias mais modernas vêm tomando destaque quando se trata de sistemas de auxílio ao diagnóstico clínico da malária. Em (Barros *et al.*, 2010), os autores utilizaram redes neurais artificiais (RNA) e redes bayesianas (RB) para o diagnóstico de malária assintomática e compararam os resultados de ambas as técnicas entre si. Nesse contexto, a técnica de redes neurais foi escolhida devido a sua capacidade de aprender a classificar padrões sem categorias bem estabelecidas, além de sua boa robustez contra o ruído nas amostras, o que fornece um método de reconhecimento de padrões eficiente. A rede Bayesiana foi utilizada devido a sua capacidade de tomar decisões, mesmo quando a sua aprendizagem se baseia em um banco de dados incompletos.

Os dados foram obtidos a partir de uma pesquisa realizada em comunidades ribeirinhas do Estado de Rondônia, no sudoeste da Amazônia brasileira, com duração de cerca de um ano, com a finalidade de estudar mais profundamente as causas que levam à malária assintomática.

A base de dados continha 380 registros, com igual proporção de pacientes saudáveis e pacientes com malária assintomática. Um conjunto de sete variáveis de entrada foi utilizado: idade, sexo, número de malárias anteriores, resultado do exame interleucina 10 (IL-10), resultado do exame interferon gama (IFN- γ), tempo vivendo na área endêmica e se dormia em rede ou não.

A IL-10 é uma citocina que tem uma função reguladora sobre as respostas imunes. Em relação a malária humana, níveis elevados de IL-10 estão associados com a gravidade reduzida. Assim, os pacientes com níveis elevados de IL-10 podem apresentar menor grau de inflamação, reduzindo a intensidade dos sintomas. Já o IFN- γ é uma citoquina com atividade pró-inflamatória. Maiores níveis de IFN- γ podem estar relacionados a uma maior chance de doença sintomática. O padrão-ouro utilizado foi o Nested PCR.

Ambas as técnicas foram comparadas a microscopia da gota espessa de sangue, método mais comumente utilizado na região. Foram utilizados 80 registros de pacientes, dos quais 40 eram de pacientes com malária assintomática e 40 de indivíduos saudáveis. Os resultados são apresentados na Tabela 3.

Tabela 3 - Resultados obtidos por (Barros *et al.*, 2010)

Método	Sensibilidade	Especificidade	Acurácia
Microscopia	22,5%	100%	61,25%
Redes Neurais	67,5%	92,5%	80%
Rede Bayesiana	37,5%	97,5%	67%

Como pode ser observado na referida tabela, a especificidade da microscopia da gota espessa foi de 100 %, o que, segundo o autor, pode ser explicado porque este método gera diagnóstico positivo apenas quando o *Plasmodium* é encontrado na amostra de sangue do paciente, por conseguinte, um falso positivo é improvável neste teste. Por outro lado, tanto o diagnóstico produzido pela técnica de redes neurais (RNA) quanto pela rede bayesiana (RB) apresentaram especificidade pouco menor, de 92,5 % e 97,5 %, respectivamente.

No contexto da malária assintomática, parece ser menos perigoso ter diagnósticos falsos positivos do que falsos negativos, porque os danos que um indivíduo infectado com malária assintomática pode sofrer se ele não receber o tratamento é maior do que o dano que um indivíduo saudável pode sofrer se ele receber um falso diagnóstico da malária e for submetido a um tratamento desnecessário. Em consequência, Barros *et al.* (2010) indica que a RNA, cujos

resultados foram superiores aos obtidos pela RB, pode ser uma opção adequada para realizar o diagnóstico em regiões endêmicas onde a prevalência da infecção assintomática é alta.

Algumas metodologias encontradas na literatura foram desenvolvidas especificamente na tentativa de solucionar o problema do diagnóstico de doenças tropicais como a malária e suas complicações, como a encefalite. Estas derivam de outras já previamente conhecidas, como é o caso da teoria apresentada em (Bogich *et al.*, 2013).

Bogich *et al.* (2013) desenvolveu um método baseado na teoria de redes para avaliar possíveis causas de surtos de doenças. Esta rede conectaria surtos de doenças conhecidas usando uma variedade de propriedades, incluindo sintomas, sazonalidade e taxa de letalidade.

Teoria de redes é o estudo das relações entre entidades ("nós") e conexões entre essas entidades ("bordas" ou "arestas"). Neste estudo, os surtos são os "nós" e uma borda é criada entre quaisquer dois surtos se eles compartilham sintomas ou têm propriedades semelhantes, como taxa de letalidade (razão entre o número de indivíduos que morrem num período específico após diagnóstico e o número de indivíduos com a doença) ou sazonalidade.

Além disso, é dada uma vantagem maior em peso se os dois focos em cada extremidade são semelhantes. Grupos de surtos que estão mais fortemente ligados uns aos outros na rede formam um *cluster*. Assim, os surtos não identificados podem se vincular a outros surtos já conhecidos, que possuam propriedades semelhantes.

O método foi testado com 125 relatos de surtos de 10 doenças infecciosas que podem causar encefalite no Sul da Ásia. Segundo (Sejvar, 2006), encefalites são infecções agudas do cérebro, muitas vezes como complicação de doenças infecciosas como sarampo, caxumba, poliomielite, hidrofobia, rubéola, catapora, raiva, AIDS e outras. Acontecem também por infecção viral, bacteriana ou por infestações de parasitas e protozoárias, como toxoplasmose ou malária. Os vírus causam inflamação do cérebro e o tecido cerebral fica inchado, o que pode destruir as células nervosas e causar sangramentos e outros danos cerebrais.

Especificamente no sul da Ásia, dez doenças são conhecidas por causarem encefalite. São elas: chandipura, febre *chikungunya*, dengue, flavivírus, malária, sarampo, meningite

asséptica, meningite bacteriana, vírus Nipah e febre tifoide. Bogich *et al.* (2013) mostraram que essas diferentes doenças frequentemente formam *clusters* distintos dentro das redes.

Em seguida, foram analisados 97 surtos de encefalite no Sul da Ásia que foram relatados sem um diagnóstico definitivo, ou seja, sem uma causa definida. Cada um deles foi associado com uma das 10 doenças, com base na qual são mais fortemente vinculados. Assim puderam identificar a doença associada mais frequentemente com um determinado surto não diagnosticado e considerar essa associação como sendo um diagnóstico primário.

O ideal é que cada grupo de surtos seja composto de uma única doença. No entanto, dados conjuntos sobrepostos de sintomas, taxa de letalidade ou sazonalidade, a maioria dos *clusters* incluía surtos de mais de uma doença, e todos eles foram considerados diagnósticos plausíveis, aumentando assim a sensibilidade, mas reduzindo a especificidade do método.

Como se pode observar na Figura 4, surtos da mesma doença (mesma cor) tendem a se agrupar. O modelo de rede atua para minimizar o número de bordas entre os surtos em diferentes comunidades da doença e maximizar o número de bordas entre focos dentro de uma única comunidade da doença. Cada círculo, chamado de "nó", representa um relatório de um surto.

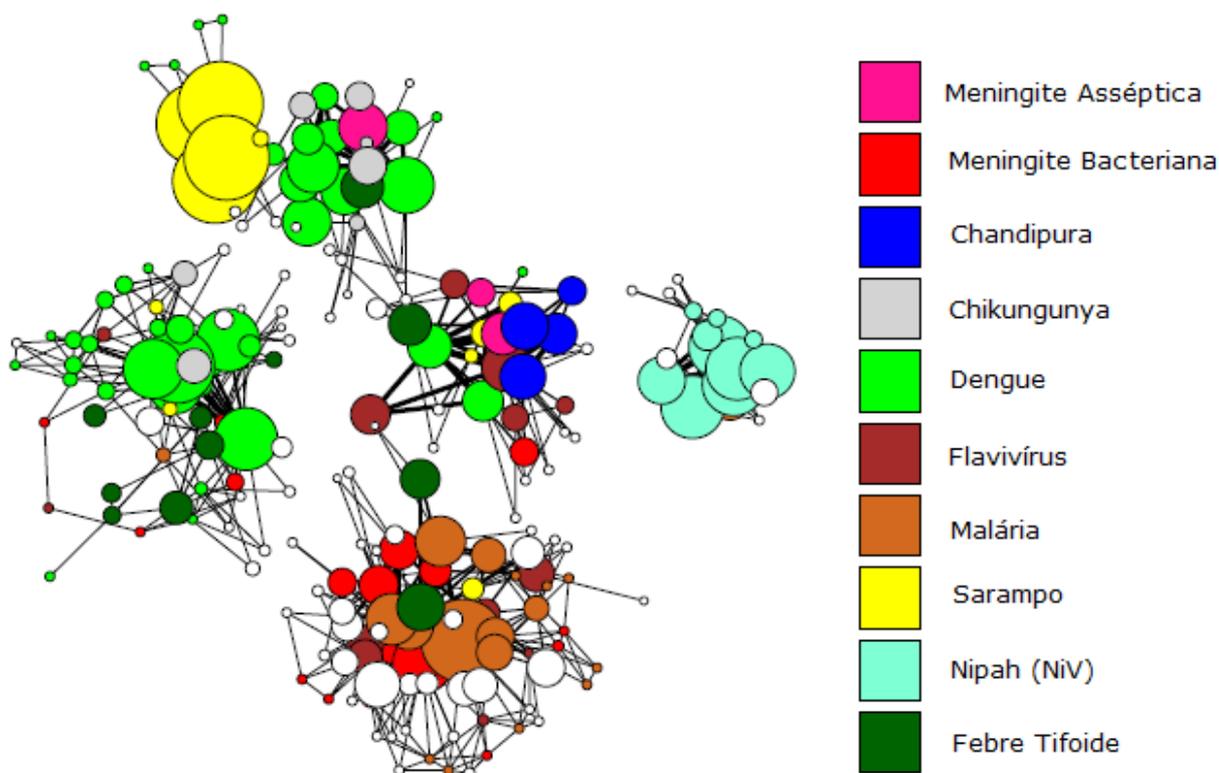


Figura 4 - Visualização da rede de surtos de doenças diagnosticadas com o potencial de causar encefalite (colorido) e surtos de encefalite não diagnosticada (branco). Fonte: Adaptado de (Bogich *et al.*, 2013)

Linhas entre dois nós indica traços compartilhados nos sintomas relatados, na taxa de letalidade ou na sazonalidade. Linhas mais grossas representam um maior número de características compartilhadas e linhas mais finas indicam menos traços compartilhados. Onde os nós se sobrepõem significa que estão fortemente ligados. O tamanho de um nó (círculo) representando um surto é proporcional à quantidade de informação contida no relatório do surto.

A malária foi identificada como doença moderadamente difícil de identificar. Foi reportado pelo autor um Valor Preditivo Positivo (proporção de pacientes com resultados verdadeiramente positivos entre os diagnosticados como positivos) de 47% com base na taxa de letalidade (aprox. 30%) e nos sintomas de perda de memória, icterícia, insuficiência renal aguda, convulsões, dificuldade respiratória e rigidez do pescoço. Dos 97 surtos não identificados, o modelo avaliou 27 como exclusivamente associados a uma única doença; 38 foram associados a duas doenças e 16 foram associados a três das 10 doenças.

Embora existam limitações para esta abordagem, este estudo fornece uma prova, de princípio, de um bom método. A precisão do método baseia-se criticamente na exatidão dos dados declarados e na integridade dos relatórios. A rede é limitada pela lista de doenças de referência (apenas 10), mas se, por exemplo, um surto não se associar a nenhuma doença conhecida no modelo, então isto seria prova para uma potencial nova doença para a região, a qual poderia ser priorizada para posterior investigação.

Uma grande vantagem desta abordagem é que ela não exige parecer de peritos ou de análise laboratorial e fornece uma forma de avaliar surtos de forma rápida e barata. Mas o autor do estudo elucida que uma comparação do método com outras técnicas de agrupamento se faz necessária.

Outras metodologias encontradas na literatura são *Fuzzy Logic* (Lógica Nebulosa) e *Analytic Hierarchy Process* (AHP).

2.3.1.1 As metodologias *Fuzzy Logic* e *Analytic Hierarchy Process* (AHP)

Em (Uzoka, Osuji e Obot, 2011), as duas metodologias, *Fuzzy* e AHP são comparadas com o diagnóstico da malária realizado pelo médico especialista. Os autores usaram uma base de dados constituída de trinta pacientes nigerianos com diagnóstico confirmado de malária. O objetivo foi estratificar a gravidade da malária em 5 graus. A casuística utilizada foi a seguinte: grau 1 – nenhum paciente; grau 2 – 2 pacientes; grau 3 – 13 pacientes; grau 4 – 13 pacientes e grau 5 – 2 pacientes.

Em resumo, as avaliações dos sintomas na folha de diagnóstico para cada paciente foram processadas usando as metodologias de AHP e *Fuzzy* de forma independente, as quais foram então comparados com o diagnóstico final do especialista, a fim de avaliar a qualidade do diagnóstico feito pelas referidas metodologias.

A seguir, será abordada uma breve conceituação dos métodos e posteriormente a descrição do caminho que os autores percorreram na utilização desses métodos.

Analytic Hierarchy Process (AHP) é um método de análise de decisão multi-critério (MDCA, do inglês *Multiple-Criteria Decision Aid*) que emprega algoritmos matemáticos para transformar julgamentos subjetivos qualitativos em dados quantitativos (Uzoka *et al.*, 2011).

O método AHP, de nome traduzido para o português como Processo Analítico Hierárquico, foi desenvolvido na década de 1970, nos Estados Unidos, por Thomas L. Saaty e seus colaboradores. Consiste na decomposição do problema por hierarquias, seguida da síntese pela identificação das relações entre as ideias. Uma hierarquia é definida como “uma abstração da estrutura de um sistema para estudar as interações funcionais de seus componentes e seus impactos no sistema total” (Saaty, 1988).

A ideia que originou o método reflete a forma pela qual a mente humana usualmente reage a um problema complexo. Diante de um grande número de elementos a serem

considerados, ela tende a dividi-los em grupos, segundo propriedades comuns, para hierarquizá-los, decompondo a complexidade encontrada, descobrindo relações entre eles e tomando a sintetizá-los (SCHMIDT, 1995).

Segundo (Saaty, 1988), quando pensamos, identificamos objetos ou ideias e também sua inter-relação. Quando identificamos alguma coisa, decompomos a complexidade encontrada. Quando descobrimos relações, sintetizamos. Esse é o processo fundamental da percepção e também da metodologia AHP: decomposição, medição de preferências e síntese.

O problema é analisado através da construção de níveis hierárquicos, ou seja, para se ter uma visão global de uma realidade complexa, o problema é decomposto em fatores. Esses fatores são decompostos em um novo nível de fatores, e assim por diante. Os elementos, previamente selecionados, são organizados numa hierarquia descendente, ficando os objetivos finais no topo e os vários resultados possíveis no nível mais baixo. O AHP parte do geral para o mais particular e concreto (SCHMIDT, 1995).

Na decomposição, o problema é quebrado em elementos gerenciáveis que são tratados individualmente. Começando com descrições implícitas do problema (o objetivo) e prosseguindo logicamente para os critérios (ou estados da natureza) em termos dos quais os resultados são avaliados. O resultado desta etapa é uma estrutura hierárquica que consiste em níveis de agrupamento de problemas (sintomas de malária) em conjunto como a importância ou influência no que diz respeito aos elementos do nível adjacente acima (Uzoka, Osuji e Obot, 2011).

A decomposição das variáveis/sintomas do diagnóstico da malária em hierarquia realizada por (Uzoka, Osuji e Obot, 2011) é apresentada na Tabela 4.

Tabela 4 – Hierarquia dos critérios básicos do diagnóstico da malária. Fonte: Adaptado de (Uzoka, Osuji e Obot, 2011)

Nível 1 (Objetivo)	Nível 2 (critérios)	Nível 3 (variáveis)
Diagnóstico da Malária	Febre (FVR)	Febre (FVR), Sudorese (EXT), Tremores (SHV)
	Dores (ACH)	Dor de cabeça (HDC), Dor muscular (MSH), Dor nas costas (BKH), Dor nas juntas (JTP)
	Sistema Nervoso Central (CNS)	Calafrios (CHL), Náuseas (NSA), Delírios (DRM), Cansaço (TRD), Sonolência (EXP), Tontura (DZN)
	Trato Gastrointestinal (GIT)	Vômitos (VMT), Diarreia (DRH), Desidratação (DRT), Desconforto Estomacal (SMC)
	Sistema Respiratório (RSS)	Respiração Anormal (ABT), Tosse (CUG)
	Males em geral (GML)	Perda de Apetite (LOA), Olhos Amarelados (YOE), Mal-estar (SUW)

A próxima etapa é a medição das preferências. Ela envolve uma comparação par a par de variáveis de avaliação (sintomas da malária). Comparações pareadas são feitas com julgamentos previstos como declarações verbais sobre a força de dominação (importância) de um sintoma em detrimento de outro sintoma, representado numericamente em uma escala absoluta.. A comparação é feita a partir do nível mais alto da hierarquia até o nível inferior, a fim de estabelecer o índice geral de prioridade. Uma escala de proporção relativa de medição é derivada a partir de comparações emparelhadas dos sintomas em um nível de hierarquia no que diz respeito à influência de um elemento no nível acima (Uzoka, Osuji e Obot, 2011).

.Ao realizar a comparação par a par, os julgamentos são expressos verbalmente como iguais, moderado, forte, muito forte, e extremo. Com esses julgamentos estão associados os números absolutos: 1, 3, 5, 7 e 9, respectivamente, os quais podem ser interpretados como, por exemplo, quantas vezes a variável A é mais importante do que a variável B). Os números 2, 4, 6, e 8 são utilizados para o compromisso entre o juízo verbal. De forma similar, os recíprocos são usados para representar a comparação inversa. Por exemplo, se A é **fortemente** mais importante do que B, então A é 5 vezes mais importante que B, então B é 1/5 vezes mais importante que A (Saaty, 1988).

As comparações, por meio das quais cada critério é confrontado com todos os outros, assumem a forma de matrizes quadradas, conforme o modelo a seguir (Almeida, 2002).

$$A = \begin{bmatrix} 1 & a_{12} & \dots & a_{1n} \\ 1/a_{12} & 1 & & a_{2n} \\ \vdots & & \ddots & \vdots \\ 1/a_{1n} & 1/a_{2n} & \dots & 1 \end{bmatrix} \quad (1)$$

As matrizes devem atender às seguintes condições:

$a_{ij} > 0$ (todos os elementos positivos)

$a_{ii} = 1$ (todos os elementos da diagonal principal iguais a 1, por comparação entre iguais)

$a_{ij} = 1/a_{ji}$ (propriedade das matrizes recíprocas)

O número de julgamentos necessários para a composição de cada matriz é $n(n - 1) / 2$, onde n é o número de critérios, coincidente com o de linhas e o de colunas.

(Uzoka, Osuji e Obot, 2011) elucidam que as matrizes de comparação em pares (*Pairwise Comparison - PWC*) foram obtidas com a ajuda de vários médicos (valores médios arredondados para números inteiros). A matriz de comparação em pares para o nível 2 da AHP está representado na Figura 5, enquanto a matriz PWC de nível 3 é mostrada na Figura 6.

	FVR	ACH	CNS	GIT	RSS	GML
FVR	1.00	5.00	7.00	7.00	8.00	6.00
ACH		1.00	7.00	6.00	7.00	3.00
CNS			1.00	1.00	3.00	0.20
GIT				1.00	6.00	0.50
RSS					1.00	0.20
GML						1.00

Figura 5- PWC do nível 2 (Uzoka, Osuji e Obot, 2011).

	FVR	EXT	SHV	HDC	MSH	BKH	JTP	CHL	NSA	DRM	TRD	EXP	DZN	VMT	DRH	DRT	SMC	ABT	CUG	LOA	YOE	GML
FVR	1.00	7.00	5.00	3.000	3.000	5.000	5.000	2.000	5.000	3.000	1.000	6.000	4.000	7.000	7.000	7.000	6.000	7.000	7.000	6.000	7.000	8.000
EXT		1.00	3.00	0.143	0.167	0.250	0.200	0.167	0.167	0.200	0.143	1.000	0.250	0.167	0.143	0.167	0.333	0.200	0.250	0.143	2.000	0.200
SHV			1.00	0.143	0.167	0.250	0.250	0.200	0.200	1.000	0.143	5.000	3.000	5.000	0.333	4.000	3.000	6.000	6.000	0.200	7.000	2.000
HDC				1.00	6.00	6.00	6.00	5.000	5.000	6.000	1.000	5.000	3.000	5.000	4.000	6.000	7.000	7.000	7.000	0.167	7.000	1.000
MSH					1.00	1.00	1.00	0.200	0.167	0.333	0.167	3.000	0.143	0.333	0.167	0.200	2.000	5.000	6.000	0.143	6.000	0.167
BKH						1.00	1.00	0.167	0.167	0.333	0.143	4.000	0.167	0.250	0.143	0.167	3.000	2.000	0.200	0.167	7.000	0.143
JTP							1.00	0.125	5.000	0.250	0.125	2.000	0.143	0.200	0.125	0.143	4.000	3.000	2.000	0.125	6.000	0.143
CHL								1.00	0.50	5.00	2.00	4.00	5.00	5.000	6.000	5.000	7.000	7.000	6.000	4.000	7.000	3.000
NSA									1.00	6.00	2.00	5.00	2.00	6.000	0.250	0.250	5.000	5.000	8.000	0.200	7.000	0.500
DRM										1.00	0.14	0.50	0.20	6.000	0.200	0.250	0.500	4.000	0.500	0.167	5.000	0.143
TRD											1.00	7.00	5.00	6.000	4.000	3.000	7.000	6.000	5.000	1.000	7.000	0.500
EXP												1.00	0.17	0.200	0.143	0.143	0.200	0.333	0.250	0.143	0.333	0.167
DZN													1.00	0.250	0.200	0.200	0.500	5.000	5.000	0.143	4.000	0.200
VMT														1.00	0.17	0.33	3.00	7.000	5.000	0.200	7.000	0.200
DRH															1.00	4.00	6.00	8.000	6.000	4.000	7.000	6.000
DRT																1.00	6.00	6.000	7.000	0.167	7.000	7.000
SMC																	1.00	3.000	0.200	0.143	2.000	0.200
ABT																		1.00	0.17	0.143	2.000	0.200
CUG																			1.00	0.143	5.000	3.000
LOA																				1.00	6.00	5.00
YOE																					1.00	0.33
GML																						1.00

Figura 6 - PWC do nível 3 (Uzoka, Osuji e Obot, 2011).

Uma vez obtidos os autovetores expressando a importância relativa dos critérios, e também dos subcritérios se for o caso, aplica-se um processo inteiramente análogo para se estabelecerem as preferências pelas alternativas em cada critério. Tendo-se as importâncias relativas dos critérios e os níveis de preferência das alternativas, efetua-se a valoração global, segundo o método da soma ponderada, pela equação:

$$V(a) = \sum_{j=1}^n p_j v_j(a) \quad (2)$$

Sujeita às condições

$$\sum_{j=1}^n p_j = 1 \quad (3)$$

$$0 < p_j < 1 \quad (j = 1, \dots, n)$$

Onde:

$V(a)$ - valor global da estimativa analisada;

p_j - importância relativa do critério j ;

v_j - nível de preferência da alternativa analisada no critério.

A última etapa do AHP é chamada de Síntese. Ela envolve o cálculo de autovalores e o autovetor. Os autovalores e o autovetor apresentam um meio de obter relações lineares entre as variáveis de avaliação e estabelecer o modelo de prioridade. É importante notar que as comparações de pares, também são realizadas para elementos dos subcritérios (variáveis) de

todos os critérios de avaliação (fatores) (Uzoka, Osuji e Obot, 2011). O autovetor é obtido como se segue:

- (a) Obter a soma da coluna;
- (b) Dividir os elementos da matriz PWC pela soma da coluna; e
- (c) Calcular a média dos elementos da linha correspondente.

Segundo Uzoka, Osuji e Obot (2011), o nível 2 gera um autovetor, K1, enquanto que as variáveis do nível 3 geram o autovetor K2. K1 combina com o vetor coluna do nível 2 para gerar o Índice de Fator de Diagnóstico para o nível 2 (DFI1), enquanto K2 combina com o vetor coluna das variáveis do nível 3 para gerar o Índice de Fator de Diagnóstico para o nível 3 (DFI2), como mostrado em (4) e (5).

$$DFI1 = 0.506FVR + 0.238ACH + 0.048CNS + 0.065GIT + 0.026RSS + 0.116GML \quad (4)$$

$$DFI2 (FVR) = 0.731FVR + 0.188SHV + 0.081EXT$$

$$DFI2 (ACH) = 0.667HDC + 0.111MSH + 0.111BKH + 0.111JTP$$

$$DFI2 (CNS) = 0.272CHL + 0.293NSA + 0.032DRM + 0.244TRD + 0.043EXP + 0.117DZN \quad (5)$$

$$DFI2 (GIT) = 0.106VMT + 0.596DRH + 0.244DRT + 0.054SMC$$

$$DFI2 (RSS) = 0.143ABT + 0.857CUG$$

$$DFI2 (GML) = 0.717LOA + 0.195YOE + 0.088SUW$$

Segundo os autores, combinando DFI1 e DFI2 teremos o Índice de Fator de Diagnóstico Agregado (ADFI), dados como:

$$\begin{aligned} ADFI = & 0.369886FVR + 0.095128SHV + 0.040986EXT + 0.158746HDC + \\ & 0.026418MSH + 0.026418BKH + 0.026418JTP + 0.013056CHL + 0.014064NSA + \\ & 0.001536DRM + 0.011712TRD = 0.002064EXP + 0.005616DZN + 0.00689VMT + \\ & 0.03874DRH + 0.01586DRT + 0.00351SMC + 0.003718ABT + 0.022282CUG + 0.083172LOA \\ & + 0.02262YOE + 0.010208SUW \end{aligned}$$

A representação gráfica de DFI1 e DFI2 é apresentada nas Figuras 7 e 8, respectivamente.

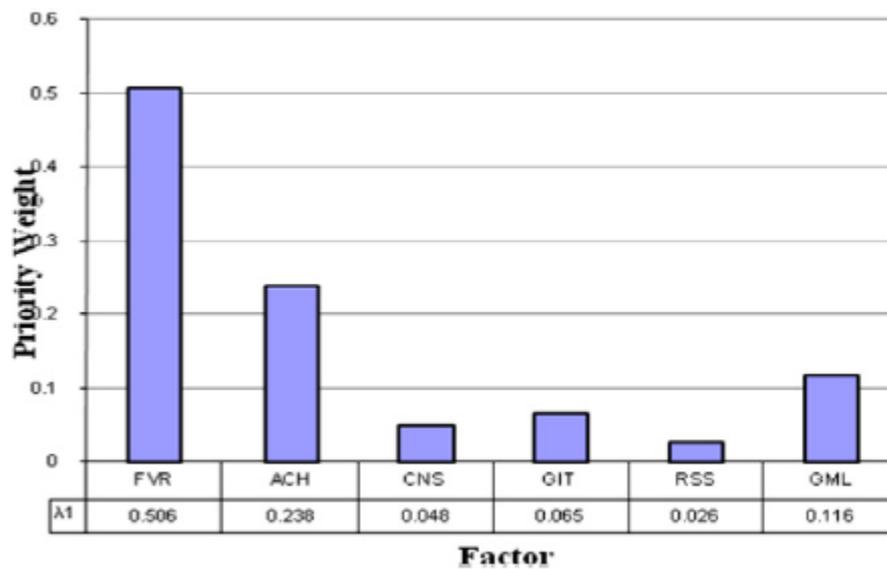


Figura 7 - DFII Prioridades do nível 2(Uzoka, Osuji e Obot, 2011)

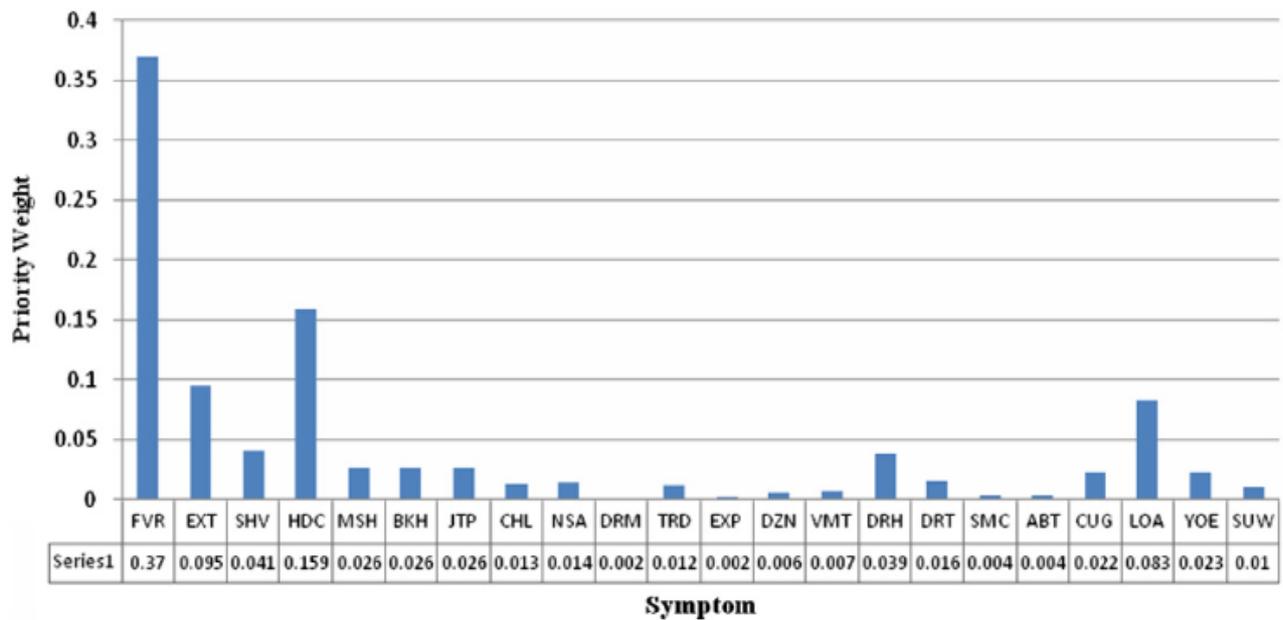


Figura 8 - DFII2 Prioridades do nível 3(Uzoka, Osuji e Obot, 2011)

Na Figura 8, observa-se que a probabilidade de um ataque de malária é mais influenciada pelas variáveis que têm valores elevados no autovetor, tais como febre (FVR) (0,37), cefaleia (HDC) (0,159), sudorese excessiva (EXT) (0,095) e perda de apetite (LOA) (0,083). Os resultados apresentados na Figura 7 também mostram que a malária tem impacto razoavelmente baixo sobre os sistemas nervoso central, respiratório e gastro intestinal.

A ADFI forma a base do diagnóstico de pacientes e determinação da intensidade de ataque de malária. A fim de determinar a intensidade de ataque de malária, Uzoka, Osuji e Obot (2011), formaram uma escala de intensidade, em que cada uma das variáveis é considerada como tendo valores uniformes desenhados a partir de uma escala de Likert (Likert, 1932), para determinar a probabilidade de ataques de malária. A escala é descrita na Tabela 5.

Tabela 5 - Escala de Intensidade de Malária (Uzoka, Osuji e Obot, 2011)

Classificação	Escala ADF1	Probabilidade de Malária
1	0 – 0.999048	Muito baixa
2	0.999049 – 1.998096	Baixa
3	1.998097 – 2.997144	Moderada
4	2.997145 – 3.996192	Alta
5	3.996193 – 4.99524	Muito alta

Embora de aplicação relativamente simples, o AHP tem um sólido embasamento teórico (Pereira, 2000), fundamentado em conceitos de cálculos matriciais e estatísticos. É vantajoso pela sua facilidade de uso e a capacidade de lidar com julgamentos inconsistentes.

Uzoka, Osuji e Obot (2011) implementaram também uma ferramenta de auxílio ao diagnóstico da malária utilizando a Lógica *Fuzzy*.

A lógica *fuzzy* foi introduzida por Lotfi Asker Zadeh, em 1965, como uma teoria matemática aplicada a conceitos difusos. A partir daí, a pesquisa e a aplicação da lógica têm crescido, inclusive tendo sido aplicada na modelagem de sistema de diagnóstico médico tanto da malária, como da hepatite viral e doenças cardiovasculares (Uzoka *et al.*, 2011).

A Lógica Nebulosa ou *Fuzzy* suporta os modos de raciocínio que são aproximados, ao invés de exatos. Essa teoria tem como base a utilização de variáveis linguísticas, cujos valores não são números e, sim, palavras ou sentenças na linguagem natural ou artificial, as quais desempenham um papel importante no tratamento da imprecisão (Zadeh, 1965).

Um sistema *fuzzy* é constituído por variáveis de entrada e saída. Para cada variável, são atribuídos conjuntos *fuzzy* que as caracterizam, sendo que, para cada conjunto *fuzzy*, é criada uma função de pertinência, ou seja, quanto que um elemento pertence a um conjunto. Na teoria dos conjuntos clássicos, deve-se decidir o grau de pertinência 0 ou 1, enquanto os conjuntos *fuzzy* nos permitem escolher um valor real arbitrário entre 0 e 1 (OLIVEIRA, Amendola e NääS, 2005; Schiassi *et al.*, 2008).

Para realização da técnica dos conjuntos *fuzzy*, as variáveis de entrada precisam sofrer uma “*fuzzificação*”. Esse processo traduz as variáveis de entrada do sistema em conjuntos *fuzzy* em seus respectivos domínios, ou seja, as variáveis e suas classificações linguísticas são catalogadas e modeladas em conjuntos *fuzzy* (Bassanezi e Barros, 2006).

A base de conhecimento utilizada pelos autores tem dois componentes, a saber:

(i) Um banco de dados, que contém informações sobre o paciente. Isto inclui seu histórico médico e informações sobre as queixas atuais.

(ii) A base de regras, que define a relação entre variáveis *fuzzy* de entrada e saída.

Com o auxílio de médicos especialistas, 112 regras *fuzzy* foram extraídas a partir dos conjuntos de dados com 22 sintomas como parâmetros de premissa.

Exemplos de regras elaboradas por Uzoka, Osuji e Obot (2011) são apresentadas a seguir:

(1) Se febre = baixa e dor de cabeça = muito baixa e dor nas juntas = intensa e náusea = moderada e desconforto estomacal = baixo e tosse = baixa e perda de apetite = muito baixa e sudorese = baixa e tremores = moderado e dor muscular = moderada e dor nas costas = baixa e delírio = moderado e cansaço = moderado e sonolência = baixa e calafrios = baixo e tontura = baixa e vômito = baixo e diarreia = moderada e respiração anormal = intensa e olhos amarelados = baixo e mal-estar = moderado e desidratação = moderada

Então a malária é moderada.

(2) Se febre = intensa e dor de cabeça = moderada e dor nas juntas = moderada e náusea = moderada e desconforto estomacal = baixo e tosse = muito baixa e perda de apetite = moderada e sudorese = intensa e tremores = muito intenso e dor muscular = baixa e dor nas costas = moderada e delírio = intenso e cansaço = intenso e sonolência = moderada e calafrios = intenso e tontura = baixa e vômito = muito baixo e diarreia = moderada e respiração anormal = muito baixa e olhos amarelados = muito intenso e mal-estar = intenso e desidratação = moderada

Então a malária é intensa.

O médico, após obter do paciente as informações de queixas e sintomas, atribui valores subjetivos para os sintomas e fornece-os ao sistema. O sistema de Uzoka, Osuji e Obot (2011) *fuzzifica* os valores de acordo com a função definida como:

$$\text{Sintoma (x)} = \begin{cases} \textit{Suave} & \textit{se } 0.1 \leq x < 0.4 \\ \textit{Moderado} & \textit{se } 0.4 \leq x < 0.6 \\ \textit{Grave} & \textit{se } 0.6 \leq x < 0.9 \\ \textit{Muito grave} & \textit{se } 0.9 \leq x < 1.0 \end{cases} \quad (6)$$

Em seguida, é realizada a inferência *fuzzy*, que consiste na realização do raciocínio *fuzzy* com base no sistema de regras que relaciona as variáveis de entrada com as de saídas (OLIVEIRA, Amendola e NääS, 2005; Schiassi *et al.*, 2008). Após a inferência da ação a ser tomada, será necessário traduzir o valor linguístico da expressão verbal para a variável numérica de saída. Esta etapa é conhecida como “*defuzzificação*” (Bilobrovec, Marçal e Kovaleski, 2004).

Um esquema de todas as etapas do sistema *Fuzzy* pode ser visto na Figura 9.

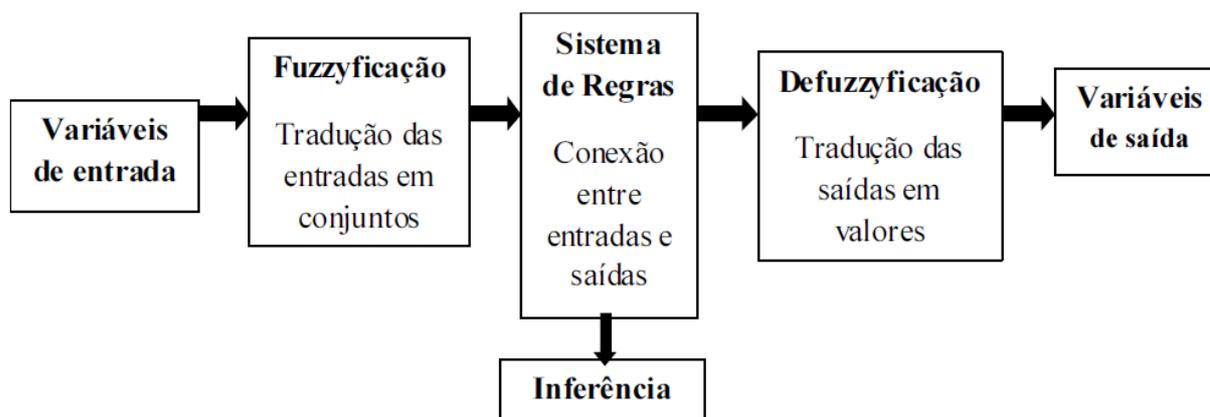


Figura 9 - Esquema ilustrativo de um sistema fuzzy. Fonte: Adaptado de (Schiassi *et al.*, 2008; Ferreira, 2009)

Na Tabela 6, são apresentados os resultados obtidos em cada método e a comparação com o diagnóstico dos médicos especialistas. A intensidade de malária foi classificada como: muito baixa (1), baixa (2), moderada (3), alta (4), e muito alta (5). Os autores reportaram uma acurácia 67% para o método AHP e de 80% para o método *fuzzy*. No total, 76,67% dos diagnósticos *fuzzy* e AHP corresponderam exatamente.

Tabela 6 - Sumário dos resultados apresentados por Uzoka, Osuji e Obot . Fonte: Adaptado de (Uzoka, Osuji e Obot, 2011)

Paciente	Diagnóstico Médico	Resultado AHP		Resultado <i>Fuzzy</i>		Diagnósticos Correspondentes
		ADFI	Escala Numérica	% Possibilidade	Escala Numérica	
1	4	3.807656	4	59	4	DAF
2	3	2.084208	3	42	3	DAF
3	4	3.112266	4	48	3	DA
4	4	3.72112	4	56	4	DAF
5	3	2.707014	3	44	3	DAF
6	3	2.905654	3	48	3	DAF
7	4	2.731552	3	54	3	AF Dev 1
8	3	2.149954	3	41	3	DAF
9	4	3.755172	4	62	4	DAF
10	3	2.511676	3	49	3	DAF
11	3	3.302706	4	54	3	DF
12	3	2.362742	3	33	2	DA
13	2	1.555794	2	27	2	DAF
14	3	2.97958	3	44	3	DAF
15	5	3.967696	4	68	4	AF Dev 1
16	4	4.248084	5	62	4	DF
17	3	2.797564	3	48	3	DAF
18	3	3.68112	4	59	4	AF Dev 1
19	4	4.15706	5	66	4	DF
20	4	3.41016	4	65	4	DAF

21	3	3.180648	4	55	3	DF
22	4	3.663516	4	62	4	DAF
23	3	2.692096	3	48	3	DAF
24	5	4.129868	5	72	5	DAF
25	4	2.846132	3	53	3	AF Dev 1
26	4	3.950988	4	61	4	DAF
27	3	2.876696	3	49	3	DAF
28	2	2.005356	3	39	2	DF
29	4	3.555852	4	63	4	DAF
30	4	3.967696	4	61	4	DAF

DAF: Os três diagnósticos são iguais. DF: O diagnóstico do médico e o *fuzzy* são iguais. DA: O diagnóstico do médico e o AHP são iguais. AF Dev 1: o diagnóstico do AHP e *Fuzzy* são iguais.

Uma análise comparativa padronizada do desempenho dos métodos foi realizada e pode ser observada na Tabela 7. Os resultados mostram que não há diferença estatística entre a AHP e lógica *fuzzy* em termos de eficácia do diagnóstico de malária. No entanto, uma inspeção rigorosa no desempenho do sistema (Tabela 6) mostra que a lógica *fuzzy* é levemente superior que a AHP na obtenção de diagnósticos de correspondência com o diagnóstico dos médicos especialistas.

Tabela 7 - Análise Comparativa das implementações AHP e *Fuzzy* dos valores dos diagnósticos normalizados

Paciente	AHP	Fuzzy
1	0.599991	0.596318
2	0.341884	0.349505
3	0.476752	0.474605
4	0.568854	0.563123
5	0.432373	0.430346
6	0.474342	0.47439
7	0.448576	0.457729
8	0.345666	0.351824
9	0.615457	0.615538
10	0.420912	0.427657
11	0.544363	0.540994
12	0.313186	0.308634
13	0.227568	0.231267
14	0.436425	0.430346
15	0.683599	0.682887
16	0.639607	0.629512
17	0.454835	0.456939
18	0.599453	0.596318
19	0.679071	0.673771
20	0.593843	0.599396
21	0.541106	0.541571
22	0.626336	0.62527
23	0.445123	0.448251
24	0.699674	0.701526

25	0.498054	0.500927
26	0.626114	0.618447
27	0.468698	0.470406
28	0.291937	0.301628
29	0.599387	0.602227
30	0.626613	0.618447

Segundo os autores, Uzoka, Osuji e Obot, (2011), as metodologias são complementares e poderiam ser usadas em combinação, pois a lógica fuzzy é capaz de lidar com incerteza e falta de estrutura na tomada de decisão, enquanto o AHP tem a capacidade de realizar a comparação de pares de elementos de decisão, a fim de determinar a sua importância no processo de decisão. Os mesmos ressaltam que os diagnósticos gerados por esses métodos se assemelham a decisão humana devido as capacidades de trabalharem com raciocínio aproximado e, encontrarem uma solução precisa.

As metodologias aqui apresentadas não são as únicas utilizadas no universo dos sistemas de auxílio ao diagnóstico médico. Muitos outros métodos foram explorados pela literatura no diagnóstico de várias outras doenças. Como por exemplo, a metodologia dos algoritmos genéticos abordados em (Berner e La Lande, 2007; Kong, Xu e Yang, 2008; Pandey e Mishra, 2009; Kasabov e Hu, 2010). Nessa revisão bibliográfica, buscou-se resumir grande parte do que tem sido feito na pesquisa de novos CDSS's, notadamente os aplicados a malária.

A Tabela 8 ao final desta seção apresenta um resumo dos artigos citados, comparando os sistemas de auxílio ao diagnóstico da malária, esclarecendo suas características e os resultados alcançados em cada um.

Tabela 8 – Principais artigos abordados na revisão bibliográfica

Ano	Autor (es)	Título	Base de Dados	Método	Diagnóstico	Características	Medida de Desempenho	Resultados			
1990	Ong Lean Suan	<i>Computer-aided Diagnosis and Treatment of Malaria: The Imex System</i>	-	Sistema Baseado em regras	Diferencial e Estadiamento	-	-	Artigo que apresenta um protótipo de sistema integrado especialista em malária			
2001	D. Chandramohan, I. Carneiro	<i>A clinical algorithm for the diagnosis of malaria: results of an evaluation in an area of low endemicity</i>	1.945 crianças febris e 2.885 adultos febris	Modelos Estatísticos	Diferencial	18 sintomas e 8 sinais	Sensibilidade, especificidade		Sens.	Esp.	
								Crianças	60%	61.2%	
								Adultos	54.6%	57.5%	
2002	D. Chandramohan, S. Jaffar and B. Greenwood	<i>Use of clinical algorithms for diagnosing malaria</i>	-	Modelos Estatísticos	Diferencial	7 sintomas e 9 sinais	Sensibilidade e Especificidade	Revisão sobre vários estudos sobre o tema			
2010	Barros Jr, Duarte	<i>Artificial Neural Networks and Bayesian Networks as Supporting Tools for diagnosis of Asymptomatic Malaria</i>	380 indivíduos (190 saudáveis e 190 com malária assintomática)	RNA e Redes Bayesianas	Diferencial	7 características (como idade, tempo vivendo em área endêmica, IL-10, IFN-gama)	Acurácia, sensibilidade e especificidade		Sens. (%)	Esp. (%)	Acurácia (%)
								RB	37.5	97.5	67%
								RNA	67.5	92.5	80
								Microscopia	22.5	100	61.25
2011	Uzoka, Osuji	<i>Clinical decision support system (DSS) in the diagnosis of malaria</i>	30 pacientes com malária diagnosticada	AHP / Fuzzy	Estadiamento	22 sintomas	Acurácia	AHP/doctor = 67%; Fuzzy/doctor = 80%; AHP/fuzzy = 76.67%			
2011	Uzoka, Osuji	<i>A Framework for Cell Phone Based Diagnosis and Management of Priority Tropical Diseases</i>	150 pacientes com doenças tropicais	AHP / Fuzzy	Diferencial	27 sintomas	-	-			
2013	Tiffany L. Bogich, Sebastian Funk, Trent R. Malcolm	<i>Using network theory to identify the causes of disease outbreaks of unknown origin</i>	125 relatos de 10 doenças que causam encefalite	Métodos dos Nós	Diferencial	10 doenças, localização geográfica, o mês e o ano do primeiro caso reconhecido, o número de pessoas afetadas, o número de mortes e sintomas clínicos.	Sensibilidade e Especificidade, PPV e NPV	Melhor resultado: Vírus Nipah: sensibilidade de 100% e 80% de PPV			

PPV = Positive Predictive Value; NPV = Negative Predictive Value; RB = Rede Bayesiana; RNA = Redes Neurais Artificiais, Sens. = Sensibilidade, Esp. = Especificidade

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo será abordada de forma breve a fundamentação teórica que subsidiará o desenvolvimento do trabalho de dissertação, quais sejam: as redes neurais artificiais (conceito, treinamento supervisionado e estratégias contra o *overfitting*: Regularização e Parada Antecipada), uma rápida introdução às medidas de separabilidade de classes e *Fisher's Discriminant Ratio* (FDR) e por fim, será feita uma revisão conceitual das técnicas de extração de características (*Principal Component Analysis* e *Linear Discriminant Analysis*).

3.1 REDES NEURAIAS ARTIFICIAIS

As redes neurais artificiais (RNA) são sistemas de processamento paralelo que tem o seu funcionamento inspirado na estrutura física do cérebro humano.

Muitos são os benefícios do uso de Redes Neurais na computação, dentre eles pode-se citar três (Tubb, 1993):

- Capacidade de aprendizagem (através de exemplos): RNA's aprendem com a experiência, objetivando aperfeiçoar seu desempenho e se adaptar aos cenários novos e dinâmicos (generalização);
- Robustez: habilidade em tratar com ruídos. São tolerantes à falhas e podem apresentar degradação gradual, ou seja, apesar de existir alguma falha no sistema, elas continuam oferecendo respostas adequadas por um período significativo de tempo;
- Velocidade de processamento: por consistirem em um grande número de unidades de processamento operando de forma paralela, elas podem operar em velocidades consideráveis em relação aos métodos computacionais comuns.

É uma ferramenta computacional muito utilizada para a solução de problemas complexos em aplicações do mundo real. Em muitos casos o desempenho das RNA's tem sido considerado superior aos métodos estatísticos (Subramanian, Hung e Hu, 1993; Falas, 1995).

Um neurônio artificial pode ser entendido como uma unidade de processamento matematicamente simples, a qual recebe uma ou mais entradas, transformando-as em saídas. Cada entrada tem um peso associado, que determina sua intensidade (Tubb, 1993). O modelo do neurônio artificial pode ser visualizado na Figura 10.

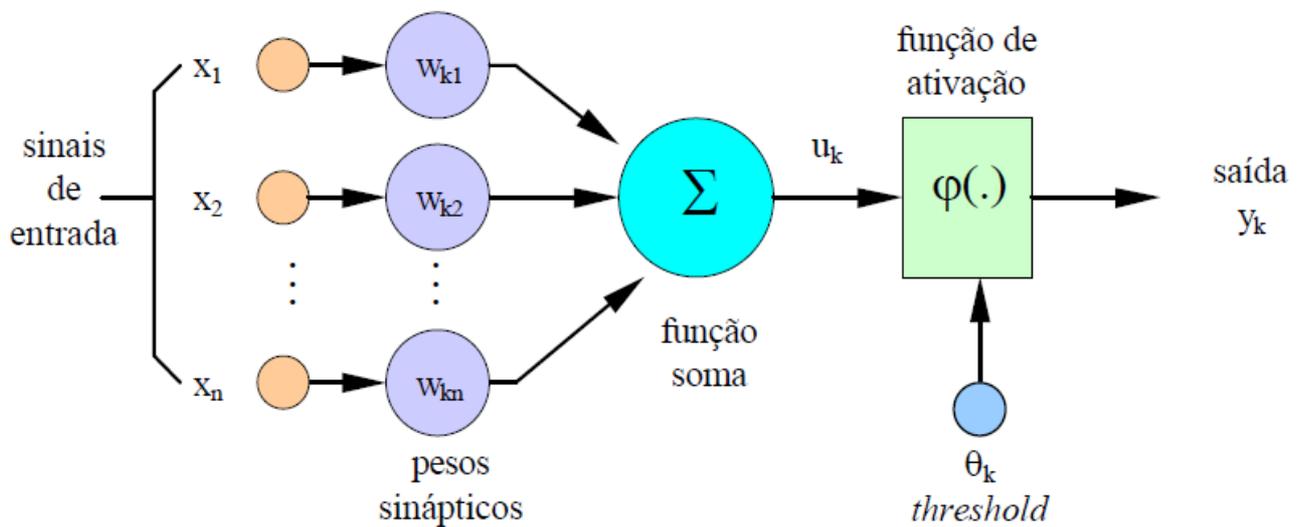


Figura 10 - Modelo não linear de um neurônio. Fonte: Adaptado de (Haykin e Network, 2004)

Os neurônios que recebem estímulos do exterior são chamados neurônios de entrada, os que se comunicam com o meio externo são chamados neurônios de saída e os demais são conhecidos como neurônios intermediários ou ocultos (*hidden*). A camada intermediária confere à rede a capacidade de resolver problemas que não são linearmente separáveis (Haykin e Network, 2004).

A aprendizagem, ou treinamento, é um processo iterativo pelo qual os pesos W_k (as ponderações) das conexões da rede neural são ajustados. A maneira particular de ajustar os pesos define o tipo específico de aprendizagem. Os métodos de aprendizagem podem ser divididos em dois grupos: aprendizado supervisionado e aprendizado não supervisionado. O aprendizado supervisionado é assim chamado por possuir um ‘supervisor’ externo que fornece a entrada e a saída desejada da rede, além de ajustar os pesos das conexões a fim de minimizar o erro. A medida do erro é efetuada comparando-se a saída obtida com a saída desejada. O aprendizado não

supervisionado, como o próprio nome sugere, não possui um ‘supervisor’. Somente os padrões de entrada são disponíveis (Braga, Carvalho e Ludermir, 2007).

3.1.1 Aprendizado Supervisionado: O Algoritmo *Backpropagation*

O Algoritmo *backpropagation* é o algoritmo de aprendizado supervisionado mais conhecido para treinamento de redes neurais multicamadas. O mesmo utiliza pares de dados (padrão de entrada, saída desejada) para ajustar os pesos das conexões das camadas intermediárias. O treinamento ocorre em duas fases, que percorrem sentidos opostos: *forward* e *backward* (Figura 11). A fase *forward* utiliza um padrão de entrada para definir a saída da rede. Os dados seguem o sentido *forward* da entrada para a saída. A fase *backward* usa a saída desejada e a saída da rede para ajustar os pesos das conexões. A diferença entre esta saída e a saída desejada constitui o erro. A intenção do trabalho de treinamento é buscar cada vez mais diminuir o valor deste erro, por isso o valor dos pesos deve ser ajustado a cada nova iteração. Os erros seguem o sentido *backward* da saída para a entrada (Braga, Carvalho e Ludermir, 2007).

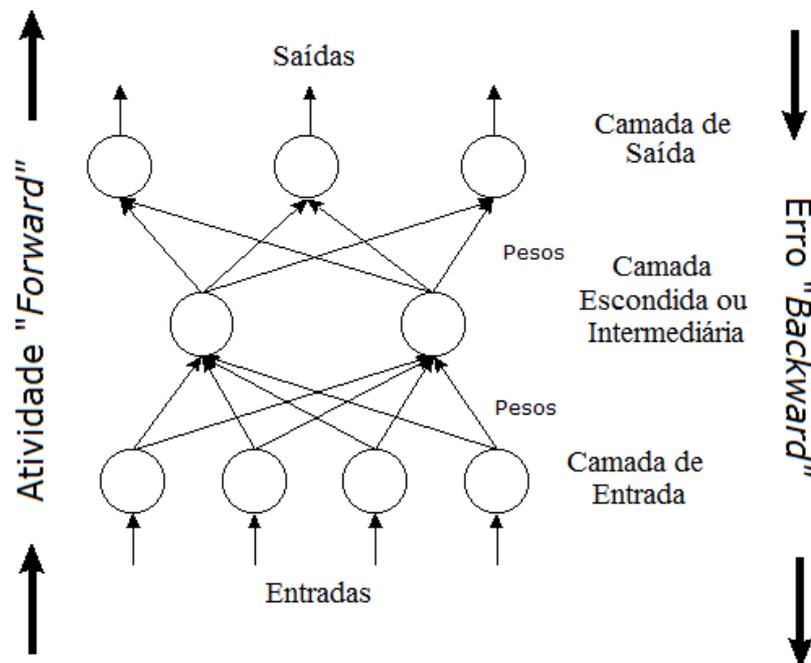


Figura 11 - Fluxo de processamento do algoritmo backpropagation. Fonte: Adaptado de (Caudill e Butler, 1992)

O algoritmo *backpropagation* é baseado na regra delta de Widrow-Hoff, e por isso pode ser chamado de regra delta generalizada. Através do algoritmo, é possível ajustar os pesos das camadas intermediárias através do método do gradiente, com o objetivo de minimizar o erro entre a saída da rede e a saída desejada (Braga, Carvalho e Ludermir, 2007).

A função de custo a ser minimizada com a derivação da regra delta generalizada é definida pela soma dos erros quadráticos, representada pela equação (7):

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (7)$$

em que:

E – medida de erro total da rede;

p – número de padrões;

k - número de unidades de saída;

d_i - a i -ésima saída desejada;

y_i - a i -ésima saída gerada pela rede.

Considerando-se que a minimização do erro para cada padrão individualmente gera a minimização do erro total, temos o erro para cada padrão definido pela equação (8).

$$E = \frac{1}{2} \sum_{j=1}^k (d_j - y_j)^2 \quad (8)$$

Segundo Braga, Carvalho e Ludermir (2007), a variação do peso, Δw_{ji} , para um dado padrão é determinada através do gradiente descendente do erro em relação ao peso, dado pela equação (9).

$$\Delta w_{ji} \propto - \frac{\partial E}{\partial w_{ji}} \quad (9)$$

Agora é necessário que se defina como cada peso deve ser ajustado para diminuir o erro total da rede. Pela regra da cadeia tem-se que:

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ji}} \quad (10)$$

em que: $net_j = \sum_{i=1}^n x_i w_{ji}$.

A derivada $\frac{\partial net_j}{\partial w_{ji}}$ calcula-se da seguinte maneira:

$$\frac{\partial net_j}{\partial w_{ji}} = \frac{\partial \sum_{i=1}^n x_i w_{ji}}{\partial w_{ji}} = x_i \quad (11)$$

A derivada $\frac{\partial E}{\partial net_j}$ mede o erro do nó j e pode ser abreviada para $\delta_j = \frac{\partial E}{\partial net_j}$. O seu cálculo também pode ser definido pela regra da cadeia:

$$\delta_j = \frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} \quad (12)$$

em que: a derivada $\frac{\partial y_j}{\partial net_j}$ pode ser assim definida:

$$\frac{\partial y_j}{\partial net_j} = \frac{\partial f(net_j)}{\partial net_j} = f'(net_j) \quad (13)$$

E o cálculo da derivada $\frac{\partial E}{\partial y_j}$, que utiliza o erro, dependerá da camada onde se encontra o nó j (Braga, Carvalho e Ludermir, 2007). Caso o nó esteja na última camada, o seu erro é definido utilizando a equação (8):

$$\frac{\partial E}{\partial y_j} = \frac{\partial (\frac{1}{2} \sum_{i=1}^k (d_i - y_i)^2)}{\partial y_j} = (d_j - y_j) \quad (14)$$

A equação (14) é a mesma que a regra delta original. Substituindo (13) e (14) na equação (12) obtêm-se:

$$\delta_j = (d_j - y_j) f'(net_j) \quad (15)$$

Se o nó j não for um nó de saída, utiliza-se a regra da cadeia para gerar:

$$\frac{\partial E}{\partial y_j} = \sum_{l=1}^M \frac{\partial E}{\partial net_l} \frac{\partial net_l}{\partial y_j} = \sum_{l=1}^M \frac{\partial E}{\partial net_l} \frac{\partial \sum_{i=1}^n w_{il} y_i}{\partial y_j} = \sum_{l=1}^M \frac{\partial E}{\partial net_l} w_{jl} \quad (16)$$

com:

$$\sum_{l=1}^M \frac{\partial E}{\partial net_l} w_{jl} = \sum_{l=1}^M \delta_l w_{jl} \quad (17)$$

Substituindo (14) e a (17) na equação (12) temos que o erro, para os nós das camadas intermediárias, é definido por:

$$\delta_j = f'(net_j) \sum_{l=1}^M \delta_l w_{jl} \quad (18)$$

Desta maneira, pode-se generalizar a equação (9), conforme segue:

$$\Delta w_{jl} = n \delta_j x_i \quad (19)$$

ou:

$$w_{ji}(t+1) = w_{ji}(t) + n\delta_j(t)x_i(t) \quad (20)$$

Caso o nó j seja de saída, o erro δ_j deve ser definido utilizando a equação (15), caso contrário utiliza-se o erro δ_j dado pela equação (18).

A atualização de pesos pode se dar de duas maneiras: por padrão ou por ciclo. Na primeira, os pesos são atualizados após a apresentação de cada padrão. Na segunda, os pesos são atualizados após a apresentação de todos os padrões. A maneira a ser utilizada depende da aplicação. Se o conjunto de dados é grande e redundante (poucos padrões e muitas entradas, por exemplo), a primeira é melhor. Caso contrário, quando o conjunto apresenta muitos padrões a serem classificados, é melhor utilizar a atualização por ciclo (Junior e Azevedo, 2000).

O algoritmo *backpropagation* apresenta alguns problemas quanto à dificuldade de aprendizado. Em superfícies muito complexas, o treinamento é muito lento e existe o risco da rede cair num mínimo local (Haykin e Network, 2004). Além disso, pode ocorrer *overfitting*, fenômeno que surge depois de certo tempo de treinamento, quando a rede memoriza padrões que já foram apresentados e o processo de classificação piora sensivelmente. A seguir apresentamos possíveis soluções para este problema.

3.2 OVERFITTING

Uma rede muito treinada pode não funcionar bem com os dados “não vistos”. A este fenômeno dá-se o nome de *overfitting* ou generalização pobre e ocorre quando uma rede neural aprende “mais do que deveria” durante o período de treinamento, perdendo assim a capacidade de generalização. Nesse caso, apesar do erro no conjunto de treinamento tenha resultado em um valor muito pequeno, quando novos dados são apresentados para a rede, o erro tende a aumentar em demasia (MacKay, 1992).

A figura a seguir adaptada de The MathWorks, Inc. (2008), mostra a resposta de uma rede neural 1-20-1 que foi treinada para se aproximar a uma função seno ruidosa. A função de seno é mostrada pela linha tracejada, as medições de ruído são dadas pelos símbolos '+', e a resposta da rede neural é dada pela linha sólida. Esse é um caso evidente que esta rede não generaliza bem.

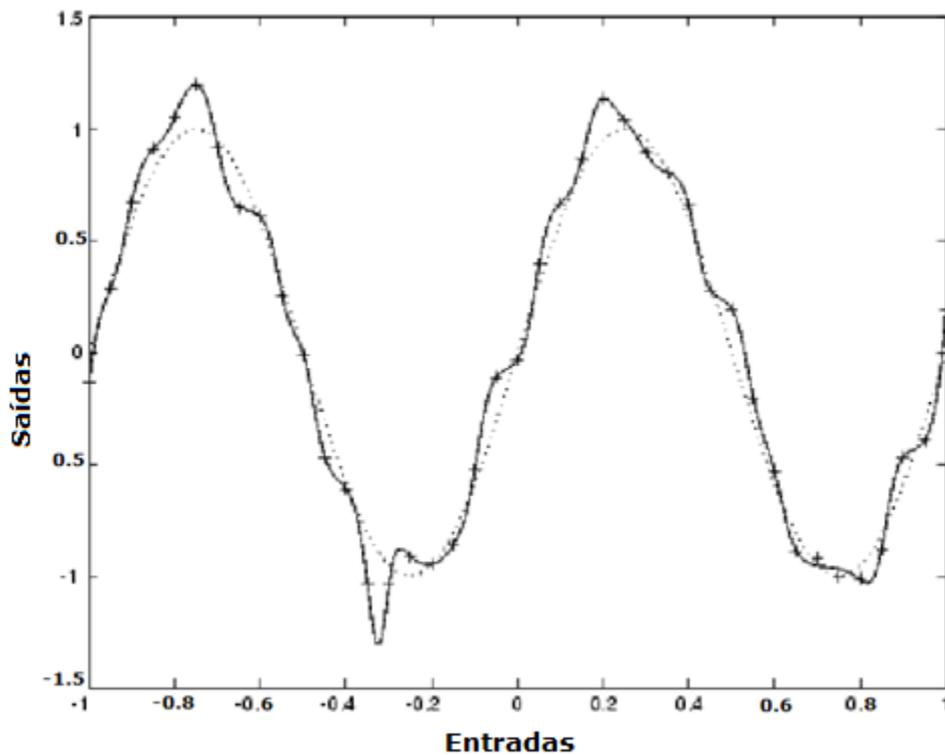


Figura 12 - Função seno com ruído representada pela linha tracejada e uma função produzida por uma rede neural 1-20-1 representada pela linha contínua. Fonte: Adaptado de The MathWorks, Inc. (2008)

Caso seja possível facilmente a obter mais dados e aumentar o tamanho do conjunto de treinamento, então não há necessidade da preocupação de se evitar *overfitting*. Isso só se aplica a situações em que se queira obter o máximo de uma oferta limitada de dados.

Várias abordagens têm sido sugeridas na literatura para solucionar esse problema. O primeiro método é um mecanismo de aprendizagem, chamado de Parada Antecipada (*Early Stopping*) em que o treinamento é concluído assim que surge *overtraining*. Isso acontece quando a precisão da previsão da rede treinada, aplicada no conjunto de teste no estágio do treinamento,

piora. A segunda abordagem é a Regularização Bayesiana. Este método minimiza o problema de *overfitting*, levando em consideração a arquitetura da rede (Doan e Liong, 2004).

3.2.1 Parada Antecipada (*Early Stopping*)

Segundo (Doan e Liong, 2004), essa abordagem requer que o conjunto de dados seja dividido em três subgrupos: treinamento, validação e teste.

O primeiro subconjunto é o conjunto de treinamento, que é usado para calcular o gradiente e atualizar os pesos da rede.

O segundo subconjunto serve para monitorar o erro durante o treinamento da rede. O erro de validação normalmente diminui durante a fase inicial de treinamento, assim como o erro do conjunto de treinamento. No entanto, quando a rede começa a “superajustar” (do inglês, *to overfit*) os dados, o erro no conjunto de validação tipicamente começa a subir. Quando o erro de validação aumenta por um determinado número de iterações, normalmente seis, o treinamento é interrompido, e os pesos no erro mínimo de validação são retornados. Na Figura 13, estão representados os erros de Treinamento e Validação e é possível observar o momento em que o erro de Validação aumenta.

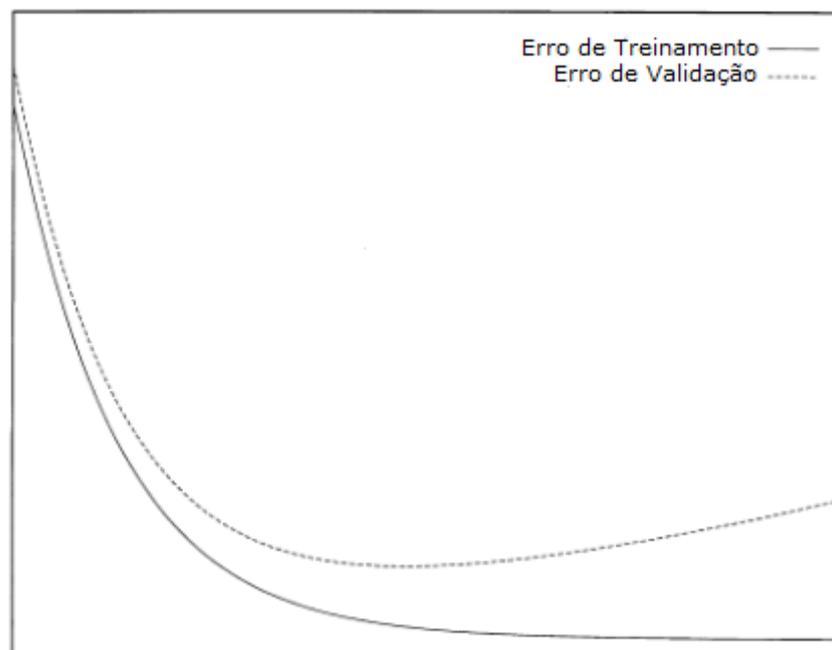


Figura 13 - Curva do Erro de Treinamento e Erro de Validação. Eixo vertical: Erro. Eixo Horizontal: Tempo. Fonte: Adaptado de (Prechelt, 1998)

O terceiro subconjunto é o conjunto de teste, utilizado para testar a acurácia do modelo treinado. Este não é usado durante o treinamento da rede, mas sim para medir desempenho e comparar com diferentes modelos. Também é útil medir o erro do conjunto de teste durante o processo de formação. Se o erro no conjunto de teste atinge um mínimo em um número de iteração significativamente diferente do que o do conjunto de validação, isto poderá indicar uma fraca divisão do conjunto de dados (Doan e Liong, 2004).

Em resumo, a principal característica deste método é que quando o erro médio quadrático do conjunto de validação cresce por seis iterações seguidas, o treinamento é finalizado. Ou seja, em fases muito posteriores do processo de treinamento, a acurácia do modelo pode começar a piorar para o conjunto de validação. Esta é a fase em que o modelo deve deixar de ser treinado para superar o problema de *overfitting*.

3.2.2 Regularização Bayesiana

Segundo Doan e Liong (2004), a regularização bayesiana faz com que o desempenho da rede tenha pesos menores, e isso força a resposta da rede ser mais suave e menos propensa a *overfitting*.. Considerando a função objetivo, F , como sendo a média dos erros quadráticos (MSE ou E_d , vide equação 21), a abordagem da Regularização Bayesiana envolve a modificação dessa função objetivo visando melhorar a capacidade de generalização do modelo.

$$F = E_d = \frac{1}{N} \sum_{i=1}^N (e_i)^2 \quad (21)$$

A função objetivo na equação (21) é expandida, com a adição de um termo E_w , que é a soma dos quadrados dos pesos da rede:

$$F = \beta E_d + \alpha E_w \quad (22)$$

Sendo α e β parâmetros da função.

O tamanho relativo dos parâmetros da função determina a ênfase para o treinamento. Se $\alpha \ll \beta$, então o algoritmo de treinamento irá conduzir a erros menores, podendo ocasionar um *overfitting*. Se um $\alpha \gg \beta$, o treinamento vai enfatizar a redução dos pesos, à custa de erros na rede, produzindo assim uma resposta de rede mais suave, mas correndo o risco da rede não se ajustar adequadamente aos dados de treinamento (Dan Foresee e Hagan, 1997).

O problema com a regularização é que é difícil determinar o valor ótimo para os parâmetros de desempenho α e β . Ainda, é desejável determinar os parâmetros ótimos de regularização de uma forma automatizada. Uma abordagem utilizada é a de David MacKay, que consiste na otimização desses parâmetros, α e β , no contexto Bayesiano (MacKay, 1992).

No aprendizado Bayesiano, os pesos da rede (vetor w) são considerados variáveis aleatórias com distribuições Gaussianas. O processo do aprendizado inicia-se com uma densidade *a priori* $p(w)$, para os parâmetros da RNA, que expresse o grau de credibilidade de cada um dos valores que os parâmetros podem possuir.

Dado um conjunto de treinamento $D = (x_1, d_1), (x_2, d_2), \dots, (x_n, d_n)$, sendo n o tamanho do conjunto de treinamento, o processo a ser modelado pode ser sua função de densidade *a priori* atualizada para *a posteriori* usando o teorema de Bayes:

$$P(w|D) = \frac{P(D|w) P(w)}{P(D)} \quad (23)$$

A densidade *a posteriori* dos parâmetros $P(w|D)$ combina a função de verossimilhança $P(D|w)$, que contém informação de w derivada do conjunto de dados, com a densidade *a priori* $P(w)$, que contém informação de w derivada de um conhecimento prévio. $P(D)$ é um fator de normalização, que garante que a probabilidade total seja 1 (MacKay, 1992).

Se assumirmos que o ruído no conjunto de treinamento é Gaussiano e a distribuição prévia dos pesos é Gaussiana, a distribuição *a priori* é escrita como uma exponencial da seguinte forma:

$$P(w) = \frac{1}{Z_W(\alpha)} \exp(-\alpha E_W) \quad (24)$$

Onde $Z_W(\alpha)$ é o fator de normalização dado por:

$$Z_W(\alpha) = \int \exp(-\alpha E_W) dw \quad (25)$$

o qual garante que a integral sobre todo o espaço dos parâmetros tenha medida 1.

Uma expressão para E_W pode ser sugerida na equação (26) no sentido de permitir que a RNA tenha uma saída suave da função que pretende-se aproximar.

$$E_w = \frac{1}{2} \|w\|^2 = \frac{1}{2} \sum_{i=1}^W w_i^2 \quad (26)$$

Esta expressão para E_w corresponde a técnica de ajuste *Weight Decay* (Moody et al., 1995) Finalmente, a distribuição *a priori* sugerida resulta em:

$$P(w) = \frac{1}{Z_W(\alpha)} \exp\left(-\frac{\alpha}{2} \|w\|^2\right) \quad (27)$$

Nota-se que um valor alto para $\|w\|$ implica em uma probabilidade *a priori* pequena para os pesos, favorecendo esta distribuição. Assume-se, até o momento, que o valor do parâmetro de regularização α é conhecido e fixo. Porém, mais adiante, será mostrado como este parâmetro é calculado como parte do processo de aprendizagem.

Uma distribuição para a função de verossimilhança geralmente utiliza uma função na forma exponencial, como a distribuição gaussiana dada pela equação (28):

$$P(D|w) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D) \quad (28)$$

onde E_D é uma função de erro, e β é um outro parâmetro de regularização do modelo neural que representa o nível de ruído da saída da rede neural. O termo $Z_D(\beta)$ presente na equação (28) é o fator de normalização expresso por:

$$Z_D(\beta) = \int \exp(-\beta E_D) dD \quad (29)$$

onde $\int dD = \int \partial d_1 \dots \partial d_N$ representa uma integral sobre todas variáveis de saída.

Assumindo que os padrões de treinamento, $D = (x_1, d_1), (x_2, d_2), \dots, (x_n, d_n)$ da função geradora a ser modelada são todos independentes, pode-se utilizar uma propriedade da estatística de

independência de eventos, pela qual a verossimilhança $P(D|w)$ pode se expressa, como produto das densidades de probabilidade da saída da RNA $p(d_n|x_n, w)$ para cada observação (parão ou amostra) das “N” observações disponíveis, da forma:

$$p(D|w) = \prod_{n=1}^N p(d_n|x_n, w) \quad (30)$$

Como a saída da RNA pode ser expressa pela seguinte equação:

$$p(d|x, w) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(d-y(x;w))^2}{2\sigma^2}\right) \quad (31)$$

onde $y(x;w)$ corresponde a saída da RNA, w representa o vetor de pesos e o parâmetro σ controla o desvio padrão do ruído. A função final da verossimilhança pode ser expressa então da seguinte maneira:

$$p(D|w) = \prod_{n=1}^N p(d_n|x_n, w) = \frac{1}{Z_D(\beta)} \exp\left(-\frac{\beta}{2} \sum_{n=1}^N \{y(x; w) - t\}^2\right) \quad (32)$$

onde $\beta = \frac{1}{\sigma^2}$. O fator de normalização $Z_D(\beta)$ é o produto das “N” integrais gaussianas independentes cuja avaliação analítica é direta, o que resulta em:

$$Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^{\frac{N}{2}} \quad (33)$$

A distribuição *a posteriori* dos parâmetros pode ser calculada por meio do teorema de Bayes da seguinte maneira:

$$p(w|D) = \frac{1}{Z_S} \exp(-\beta E_D - \alpha E_W) = \frac{1}{Z_S} \exp(-S(w)) \quad (34)$$

onde $Z_S(\alpha, \beta) = \int \exp(\beta E_D + \alpha E_W) dw$ é o fator de normalização. A função de erro $E = -\ln p(D/w)$ para este modelo neural é

$$E = S(w) = \beta E_D + \alpha E_W \quad (35)$$

O valor mais provável para o vetor de parâmetros $w = [w_1, w_2, \dots, w_W]$ é o valor que corresponde à probabilidade máxima *a posteriori* w_{MP} , ou seja, o valor mais provável para os pesos da rede neural.

Em MacKay (1992) utiliza-se uma aproximação gaussiana da distribuição *a posteriori* dos pesos, obtida por meio da expansão da série de Taylor de $S(w)$ sobre um ponto qualquer \hat{w} :

$$S(w) = S(\hat{w}) + (w - \hat{w})^T \nabla S(\hat{w}) + \frac{1}{2!} (w - \hat{w})^T \nabla \nabla S(\hat{w}) (w - \hat{w}) + \dots \quad (36)$$

Determinar o α e β ótimo para a função de custo é a grande questão do método de Regularização Bayesiana. O cálculo dos parâmetros de regularização α e β devem ser tais que maximizem a função de densidade *a posteriori* dos parâmetros $p(w/D)$ que pode ser expressa pela equação:

$$p(w/D) = \iint p(w, \alpha, \beta | D) d\alpha d\beta \quad (37)$$

A dimensão do vetor de parâmetros torna a solução analítica da integral bastante difícil. Uma solução encontrada por MacKay (1992) é uma aproximação gaussiana que torna a densidade *a posteriori* desta maneira:

$$p(w/D) = p(w | \alpha_{MP}, \beta_{MP}, D) \quad (38)$$

onde α_{MP} e β_{MP} representam os valores máximos *a posteriori* da densidade *a posteriori* $p(\alpha, \beta | D)$, expressa do seguinte modo:

$$p(\alpha, \beta | D) = \frac{p(D | \alpha, \beta) p(\alpha, \beta)}{p(D)} \quad (39)$$

Desta maneira o logaritmo natural da verossimilhança $\ln p(D | \alpha, \beta)$ é

$$\ln p(D | \alpha, \beta) = -\alpha E_D^{MP} - \beta E_W^{MP} - \frac{1}{2} \ln |H| + \frac{W}{2} \ln \alpha + \frac{N}{2} \ln \beta - \frac{N}{2} \ln (2\pi) \quad (40)$$

Nesta expressão precisa-se do valor de $\ln |H|$. A matriz Hessiana H , pode ser expressa por $H = \beta B + \alpha G$, onde $B = \nabla \nabla E_D^{MP}$ e $\nabla \nabla E_W^{MP} = I$. Se $\{\lambda_i\}$ são os autovalores de B , então H tem $\{\lambda_i + \alpha\}$ sendo $i = 1, 2, \dots, W$. Desta forma, como demonstrado em (Bishop, 1995):

$$\frac{d}{d\alpha} \ln |H| = \sum_{i=1}^W \frac{1}{\lambda_i + \alpha} \quad (41)$$

Substituindo em (40) temos:

$$2\alpha E_W^{MP} = W - \sum_{i=1}^W \frac{1}{\lambda_i + \alpha} \quad (42)$$

Seja $\gamma = \sum_{i=1}^W \frac{\lambda_i}{\lambda_i + \alpha}$, o número efetivo de parâmetros estimado para a RNA, temos a equação do cálculo dos parâmetros de regularização:

$$\alpha^{MP} = \frac{\gamma}{2E_W(w^{MP})} \quad (43)$$

$$\beta^{MP} = \frac{n - \gamma}{2E_D(w^{MP})} \quad (44)$$

A otimização Bayesiana dos parâmetros da regularização requer a obtenção da matriz Hessiana de $F(\mathbf{w})$ no ponto mínimo, \mathbf{w}^{MP} . Foresee e Hagan (1997) propõem o uso da aproximação Gauss-Newton para a matriz Hessiana, que pode ser convenientemente executada no âmbito do algoritmo de Levenberg-Marquardt, para reduzir a sobrecarga computacional.

O quadro 1 apresenta os passos para otimização Bayesiana dos parâmetros da regularização, com a aproximação Gauss-Newton para matriz Hessiana, segundo Foresse e Hagan (1997),

1) Inicializar α , β e os pesos. Pode-se iniciar, por exemplo, com $\alpha = 0$ e $\beta = 1$ e utilizar o método Nguyen-Widrow de inicialização dos pesos (Nguyen e Widrow, 1990). Depois do primeiro passo do treinamento, os parâmetros da função objetivo vão recuperar a configuração inicial.

2) Dar um passo no algoritmo Levenberg-Marquardt para minimizar a função

$$F(\mathbf{w}) = \beta E_D + \alpha E_W.$$

3) Calcular a quantidade de parâmetros $\gamma = N - 2\alpha \text{tr}(\mathbf{H})^{-1}$ usando a aproximação Gauss-Newton para a Hessiana disponível no algoritmo de treinamento Levenberg-Marquardt:

$$\mathbf{H} = \nabla^2 F(\mathbf{w}) \approx 2\beta \mathbf{J}^T \mathbf{J} + 2\alpha \mathbf{I}_N$$

em que \mathbf{J} é a matriz Jacobiana do conjunto de treinamento.

4) Calcular as novas estimativas para os parâmetros da função objetivo

$$\alpha = \frac{Y}{2E_T(w)} \text{ e } \beta = \frac{n-Y}{2E_D(w)}$$

5) Repetir os passos 2 a 4 até convergir.

Quadro 1: Algoritmo de otimização Bayesiana dos parâmetros da regularização, com a aproximação Gauss-Newton para matriz Hessiana. Fonte: Foyse e Hagan, 1997.

Na Figura 14 está ilustrado o resultado de um treinamento do algoritmo da Aproximação Gauss-Newton para Regularização Bayesiana (GNBR). Neste exemplo, 100 pontos com ruído Gaussiano de média zero e variância 0.01 foram acrescentados a pontos amostrados de uma onda triangular. O primeiro gráfico mostra uma rede 1-6-1 sem regularização e o segundo, uma rede 1-6-1 treinada com o algoritmo GNBR. É perceptível o *overfitting* no treinamento sem a regularização.

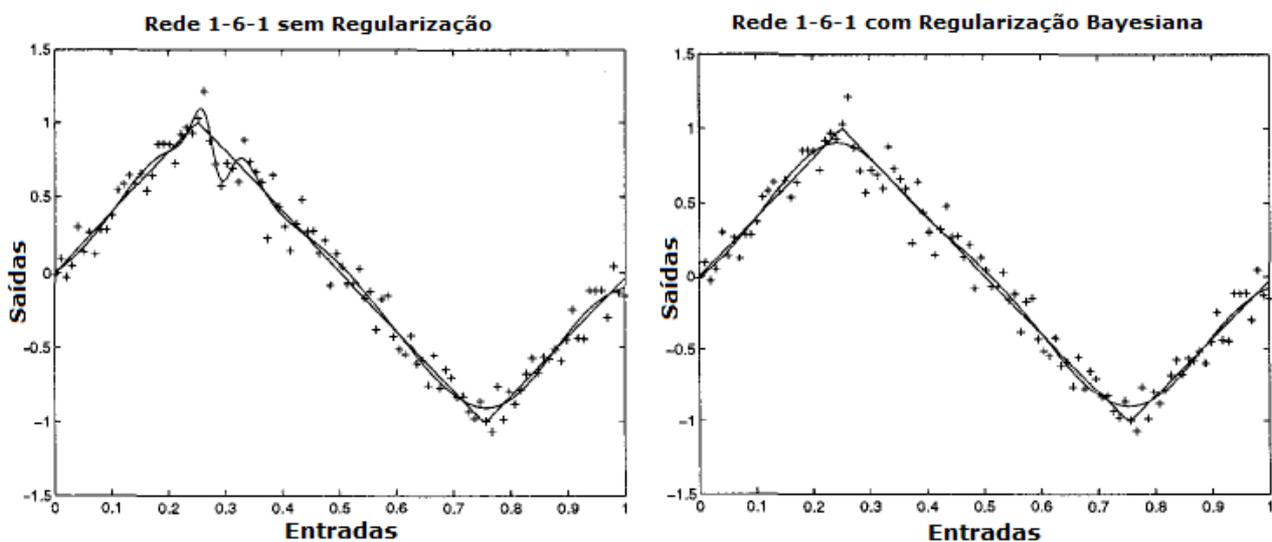


Figura 14 - Função triangular com 100 pontos de ruído (representados por +) e uma função produzida por rede neural representada pela linha curva. No lado esquerdo, a rede neural sem regularização e no lado direito, a rede neural com regularização bayesiana. Fonte: Adaptado de Foresee e Hagan (1997).

3.3 TÉCNICAS DE REDUÇÃO DE DIMENSIONALIDADE

A extração de informações dos resultados de um experimento envolve a análise de um número de variáveis. Muitas vezes, um pequeno número destas variáveis adiciona pouco ou nada à interpretação dos resultados. A decisão sobre quais variáveis são importantes é baseada, geralmente, em critérios subjetivos (Neto e Moita, 1998).

A redução de variáveis através de critérios objetivos, pode ser conseguida, por exemplo, através da análise de componentes principais (PCA) e da análise discriminante linear (LDA).

3.3.1 ANÁLISE DOS COMPONENTES PRINCIPAIS (PCA)

O PCA é um dos métodos mais populares em reconhecimento de padrões para geração de características e redução de dimensionalidade. Apesar de ser uma técnica antiga, ainda está em uso, e forma a base para outras abordagens mais avançadas. É basicamente uma técnica estatística usada para simplificar um conjunto de dados. Seus fundamentos encontram-se descritos nos artigos de seus desenvolvedores (Pearson, 1901) e (Hotelling, 1933), enquanto que a melhor referência moderna é (Jolliffe, 2002).

O objetivo do método é reduzir a dimensionalidade dos dados multivariados, preservando o máximo de informações relevantes possível. É uma forma de aprendizagem não supervisionada na medida em que se baseia inteiramente na entrada de dados em si, sem referência aos dados de destino correspondentes (o critério a ser maximizado é a variância). Em resumo, o PCA consiste em transformar um conjunto de variáveis originais em outro conjunto de variáveis denominadas de componentes principais.

Os componentes principais apresentam propriedades importantes: cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e

estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos da variação total contida nos dados. (Hotelling, 1933)

A análise de componentes principais é associada à ideia de redução de massa de dados, com menor perda possível da informação. Procura-se redistribuir a variação observada nos eixos originais de forma a se obter um conjunto de eixos ortogonais não correlacionados.

Esta técnica pode ser utilizada para geração de índices e agrupamento de indivíduos. A análise agrupa os indivíduos de uma população de acordo com a variação de suas características (Varella, 2008). Considerando ‘p’ características de ‘n’ indivíduos de uma população π . As características observadas são representadas pelas variáveis $X_1, X_2, X_3, \dots, X_p$. A matriz de dados é de ordem ‘n x p’ e normalmente denominada de matriz ‘X’.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{np} \end{bmatrix} \quad (45)$$

O objetivo da análise de componentes principais é transformar essa estrutura, representada pelas variáveis $X_1, X_2, X_3, \dots, X_p$ em uma outra estrutura representada pelas variáveis $Y_1, Y_2, Y_3, \dots, Y_p$ não correlacionadas e com variâncias ordenadas, para que seja possível comparar os indivíduos usando apenas as variáveis Y_{is} que apresentam maior variância (representada por $\hat{V}ar$) (Varella, 2008).

A partir da matriz X de dados de ordem ‘n x p’ é possível fazer uma estimativa da matriz de covariância Σ da população π , que representaremos por S. A matriz de covariância S é simétrica e de ordem ‘p x p’.

$$S = \begin{bmatrix} \hat{V}ar(x_1) & \hat{C}ov(x_1x_2) & \hat{C}ov(x_1x_3) & \dots & \hat{C}ov(x_1x_p) \\ \hat{C}ov(x_2x_1) & \hat{V}ar(x_2) & \hat{C}ov(x_2x_3) & \dots & \hat{C}ov(x_2x_p) \\ \hat{C}ov(x_3x_1) & \hat{C}ov(x_3x_2) & \hat{V}ar(x_3) & \dots & \hat{C}ov(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{C}ov(x_px_1) & \hat{C}ov(x_px_2) & \hat{C}ov(x_px_3) & \dots & \hat{V}ar(x_p) \end{bmatrix} \quad (46)$$

Normalmente as características são observadas em unidades de medidas diferentes entre si, e neste caso, segundo Regazzi (2000) é conveniente padronizar as variáveis X_j ($i = 1, 2, 3, \dots, p$).

A padronização pode ser feita com média zero e variância 1, ou com variância 1 e média qualquer:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s(x_j)}, \quad i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p \quad (47)$$

$$Z_{ij} = \frac{x_{ij}}{s(x_j)}, \quad i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p \quad (48)$$

Em que \bar{x}_j e $s(x_j)$ são, respectivamente, a estimativa da média e o desvio padrão da característica j :

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad (49)$$

$$s(x_j) = \sqrt{\hat{V}ar(x_j)} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}, \quad j = 1, 2, \dots, p \quad (50)$$

$$\text{Ou } \sqrt{\hat{V}ar(x_j)} = \frac{\sum_{i=1}^n x_{ij}^2 - \frac{(\sum_{i=1}^n x_{ij})^2}{n}}{n-1}, \quad (51)$$

Após a padronização obtemos uma nova matriz de dados R. Os componentes principais são determinados resolvendo-se a equação característica da matriz, isto é:

$$\det[R - \lambda I] = 0 \quad (52)$$

$$R = \begin{bmatrix} 1 & r(x_1x_2) & r(x_1x_3) & \dots & r(x_1x_p) \\ r(x_2x_1) & 1 & r(x_2x_3) & \dots & r(x_2x_p) \\ r(x_3x_1) & r(x_3x_2) & 1 & \dots & r(x_3x_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(x_px_1) & r(x_px_2) & r(x_px_3) & \dots & 1 \end{bmatrix} \quad (53)$$

Se a matriz R for de posto completo igual a 'p', isto é, se R não apresentar nenhuma coluna que seja combinação linear de outra, a equação $|R - \lambda I| = 0$ terá 'p' raízes chamadas de autovalores ou raízes características da matriz R. Na montagem da matriz de dados X é importante observar que o valor de 'n' (indivíduos) dever ser pelo menos igual a 'p + 1', isto é, se queremos montar um experimento para analisar o comportamento de 'p' características de indivíduos de uma população é

recomendado que o delineamento estatístico apresente pelo menos ‘p+1’ tratamentos (Regazzi, 2002).

Sejam $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ as raízes da equação característica da matriz R, então: $\lambda_1 > \lambda_2 > \lambda_3 \dots \lambda_p$. Para cada autovalor λ_i existe um autovetor \tilde{a}_i :

$$\tilde{a}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix} \quad (54)$$

Os autovetores \tilde{a}_i são normalizados, isto é, a soma dos quadrados dos coeficientes é igual a 1, e ainda são ortogonais entre si. Sendo \tilde{a}_i o autovetor correspondente ao autovalor λ_i , então o i-ésimo componente principal é dado por:

$$Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad (55)$$

Segundo Hotelling (1933), os componentes principais apresentam as seguintes propriedades:

- 1) A variância do componente principal Y_i é igual ao valor do autovalor λ_i
- 2) A maior variância por qualquer projeção dos dados se encontra na primeira coordenada, a segunda maior variância na segunda coordenada, e assim por diante.
- 3) O total de variância das variáveis originais é igual ao somatório dos autovalores que é igual ao total de variância dos componentes principais.

$$\sum \hat{V}ar(X_i) = \sum \lambda_i = \sum \hat{V}ar(Y_i) \quad (56)$$

- 4) Os componentes principais não são correlacionados entre si:

$$\hat{C}ov(Y_i, Y_j) = 0 \quad (57)$$

A importância de um componente principal é avaliada por meio de sua contribuição, isto é, pela proporção de variância total. A contribuição C_i de cada componente principal Y_i é expressa em porcentagem. É calculada dividindo-se a variância de Y_i pela variância total.

$$C_i = \frac{\hat{V}ar(Y_i)}{\sum_{i=1}^p \hat{V}ar(Y_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100 = \frac{\lambda_i}{\text{traço}(S)} \cdot 100 \quad (58)$$

A soma dos primeiros k autovalores representa a proporção de informação retida na redução de p para k dimensões. Com essa informação podemos decidir quantos componentes serão usados

na análise. Em aplicações em diversas áreas do conhecimento, os componentes utilizados tem sido aqueles que juntos acumulam 70% ou mais de proporção da variância total (Regazzi, 2002).

3.3.2 ANÁLISE DISCRIMINANTE LINEAR (LDA)

A expressão Análise Discriminante tem sido utilizada para identificar diversas técnicas multivariadas que, no entanto, têm um objetivo comum. Parte-se do conhecimento de que os n indivíduos observados pertencem a diversos subgrupos e procura-se determinar funções das p variáveis observadas que melhor permitam distinguir ou discriminar entre esses subgrupos ou classes (Koutroumbas, 2008).

Nesta seção será abordada uma técnica discriminante, conhecida por Análise Discriminante Linear, ou de Fisher. Essa técnica é empregada para redução da dimensionalidade de dados pertencentes a várias classes, procurando escolher uma direção ou um conjunto de direções no espaço que possibilitem otimizar um critério de separabilidade das classes (Koutroumbas, 2008).

A discriminação de Fisher tem a virtude de ser facilmente visualizável em termos geométricos. Além disso, não exige hipóteses adicionais (ao contrário das técnicas baseadas em modelos probabilísticos) e permite discriminar mais que dois diferentes subgrupos (classes) sem grande complexidade.

Segundo Koutroumbas (2008), na Análise Discriminante de Fisher procuram-se as combinações lineares X_{α} das p variáveis observadas que melhor separem os subgrupos de indivíduos indicados. As soluções X_{α} obtidas designam-se **eixos discriminantes** ou também **variáveis canônicas**. Podem ser utilizados para obter uma representação gráfica que saliente a distinção entre as classes, e para classificar futuros indivíduos (observados nas mesmas variáveis), do qual seja desconhecido o subgrupo a que pertence.

O objetivo da LDA é encontrar um sub-espaço, que maximize um critério de separabilidade das classes. Esse sub-espaço é denominado sub-espaço principal.

Considere um vetor \mathbf{x} no espaço m-dimensional. Assuma que seus pontos são originários de duas classes e que se deseja "espremer" a informação de classificação relacionada com \mathbf{x} em um menor número (neste caso apenas uma) de características (Koutroumbas, 2008). Isso corresponde a gerar uma característica y como uma combinação linear dos componentes de \mathbf{x} .

Este objetivo será alcançado, buscando a direção \mathbf{w} no espaço m-dimensional, de forma que as duas classes sejam separadas de alguma forma.

A projeção de \mathbf{x} em \mathbf{w} será:

$$y = \frac{\mathbf{w}^T \mathbf{x}}{\|\mathbf{w}\|} \quad (59)$$

Considerando que o escalonamento de todos os vetores de características pelo mesmo fator não adiciona qualquer informação de classificação, o fator de escala $\|\mathbf{w}\|$ será desprezado.

Em seguida, adota-se a relação discriminante de Fisher (FDR) (Kim, Magnani e Boyd, 2005):

$$FDR = \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \quad (60)$$

Em que μ_1, μ_2 são as médias e σ_1^2, σ_2^2 as variâncias de y nas duas classes ω_1, ω_2 , respectivamente, depois da projeção em \mathbf{w} .

A relação discriminante de Fisher é utilizada para quantificar a capacidade de separabilidade de características individuais dos componentes de uma classe. O maior valor da razão FDR representa a característica que melhor separa duas classes (Koutroumbas, 2008).

Usando a definição em (47) e omitindo $\|\mathbf{w}\|$, é facilmente visto que:

$$\mu_i = \mathbf{w}^T \boldsymbol{\mu}_i, \text{ sendo } i = 1, 2 \quad (61)$$

Em que $\boldsymbol{\mu}_i$ é a média dos dados em ω_i no espaço m-dimensional.

Sejam \mathbf{S}_w e \mathbf{S}_b as matrizes de covariância entre classes e interclasses e entre classes, respectivamente, dadas por:

$$\begin{aligned} S_w &= \sum_{i=1}^M P_i (x - \mu_i)(x - \mu_i)^T \\ S_b &= \sum_{i=1}^M P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \end{aligned} \quad (62)$$

Onde P_i é a probabilidade *a priori* da classe ω_i e μ_0 é o vetor médio global:

$$\mu_0 = \sum_i P_i \mu_i \quad (63)$$

Assumindo que as classes são equiprováveis, temos que:

$$(\mu_1 - \mu_2)^2 = \mathbf{w}^T (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \mathbf{w} \propto \mathbf{w}^T S_b \mathbf{w} \quad (64)$$

Onde \propto quer dizer “proporcional a”. Agora voltando a atenção ao denominador de (60), temos:

$$\sigma_i^2 = E[(y - \mu_i)^2] = E[\mathbf{w}^T (x - \mu_i)(x - \mu_i)^T \mathbf{w}] \quad (65)$$

$$\sigma_1^2 + \sigma_2^2 \propto \mathbf{w}^T S_w \mathbf{w} \quad (66)$$

Combinando (66), (64) e (60), teremos que a direção ótima é obtida maximizando o critério de Fisher com relação a \mathbf{w} :

$$FDR(\mathbf{w}) = \frac{\mathbf{w}^T S_b \mathbf{w}}{\mathbf{w}^T S_w \mathbf{w}} \quad (67)$$

$$\mathbf{w} = S_w^{-1} (\mu_1 - \mu_2) \quad (68)$$

Assumindo que S_w seja invertível.

Segundo Koutroumbas (2008), em LDA adota-se o critério J_3 de separabilidade entre classes, dado por:

$$J_3 = \text{traço} \left(\frac{S_b}{S_w} \right) \quad (69)$$

Para maximizar o critério J_3 , a matriz U é formada pelos autovetores que correspondem aos maiores autovalores da matriz $S_w^{-1} S_b$: $U = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_l]$ (dimensão $d \times l$).

Em que: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_l$ são os autovetores que correspondem aos l maiores autovalores de $S_w^{-1} S_b$. A dimensão do vetor \mathbf{a}_p é obtida a partir de $\mathbf{a}_p = U^T \mathbf{x}$: $(l \times d) (d \times 1) \rightarrow (l \times 1)$.

A redução de dimensionalidade é obtida considerando que $l < d$. Quanto menos autovetores \mathbf{u}_i forem utilizados na montagem da matriz U, maior a redução da dimensionalidade (Koutroumbas, 2008).

MATERIAIS E MÉTODOS

Este capítulo descreve as diferentes metodologias utilizadas neste trabalho. Estas foram desenvolvidas com base nos objetivos específicos que norteiam o trabalho. As metodologias serão avaliadas e comparadas a fim de que se desenvolva uma metodologia de auxílio na avaliação da gravidade da malária em pacientes com diagnóstico confirmado.

4.1 MATERIAIS

O material a ser utilizado nesse trabalho para o auxílio na avaliação da gravidade da malária em pacientes com diagnóstico confirmado será aquele disponibilizado no trabalho de Uzoka, Osuji e Obot (2011). Essa base de dados é constituída de informações clínicas de 30 pacientes com diagnóstico confirmado de malária.

Esses dados consistem de um conjunto de 22 sintomas:

- | | |
|-------------------|--------------------------|
| 1. Febre | 12. Sonolência |
| 2. Sudorese | 13. Tontura |
| 3. Tremor | 14. Vômito |
| 4. Dor de cabeça | 15. Diarreia |
| 5. Dor muscular | 16. Desidratação |
| 6. Dor nas costas | 17. Desconforto gástrico |
| 7. Dor nas juntas | 18. Respiração anormal |
| 8. Calafrios | 19. Tosse |
| 9. Náuseas | 20. Perda de apetite |
| 10. Delírio | 21. Olhos amarelados |
| 11. Cansaço | 22. Indisposição |

Os dados dos 30 pacientes foram coletados por seis médicos de quatro hospitais (dois públicos e dois privados) da Nigéria, onde médicos especialistas em malária são abundantes.

Cada médico realizou a avaliação diagnóstica de cinco pacientes com malária a fim de definir a intensidade da malária que acometia cada um dos pacientes. Além do diagnóstico médico final indicando a intensidade do ataque da malária, cada paciente teve cada um dos 22 sintomas classificado como: 1 – muito fraco, 2 - fraco, 3 – moderado, 4 – intenso ou 5 – muito intenso.

4.2 MÉTODOS

O modelo computacional utilizado no reconhecimento de padrões são as Redes Neurais Artificiais.

O desenvolvimento da proposta divide-se em três fases. As mesmas se distinguem pelo conjunto de dados utilizado como entrada das RNA's implementadas. A Figura 15 ilustra as três fases com respeito ao conjunto de dados. A primeira fase é caracterizada por utilizar o conjunto de dados originais com 22 características. A segunda fase, por utilizar extração de características com o método PCA, reduzindo de 22 sintomas, primeiramente para 15 e depois para 10 características. E finalmente, a terceira fase, análoga a segunda, porém com o auxílio do método LDA. Para fins de comparação, o conjunto de dados também será reduzido com as mesmas quantidades de componentes testadas na fase anterior.

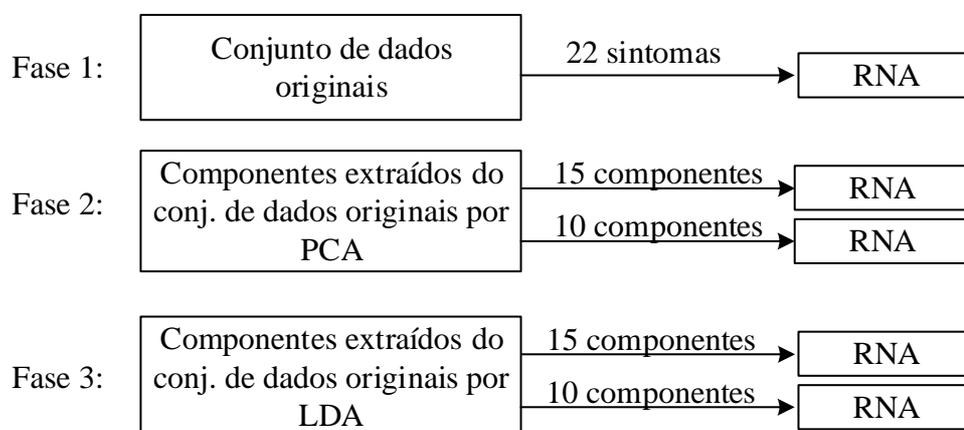


Figura 15- Três fases da pesquisa e os diferentes conjuntos de dados utilizados

No diagrama de blocos da Figura 16 estão descritas as quatro arquiteturas utilizadas nas RNA's. As arquiteturas estão representadas por $cd_i - n_{hi} - 5$. Onde cd_i é o número de neurônios na camada de entrada, n_{hi} é o número de neurônios na camada intermediária e 5 é o número de neurônios na camada de saída. Esta arquitetura é discutida na sessão 4.2.4.

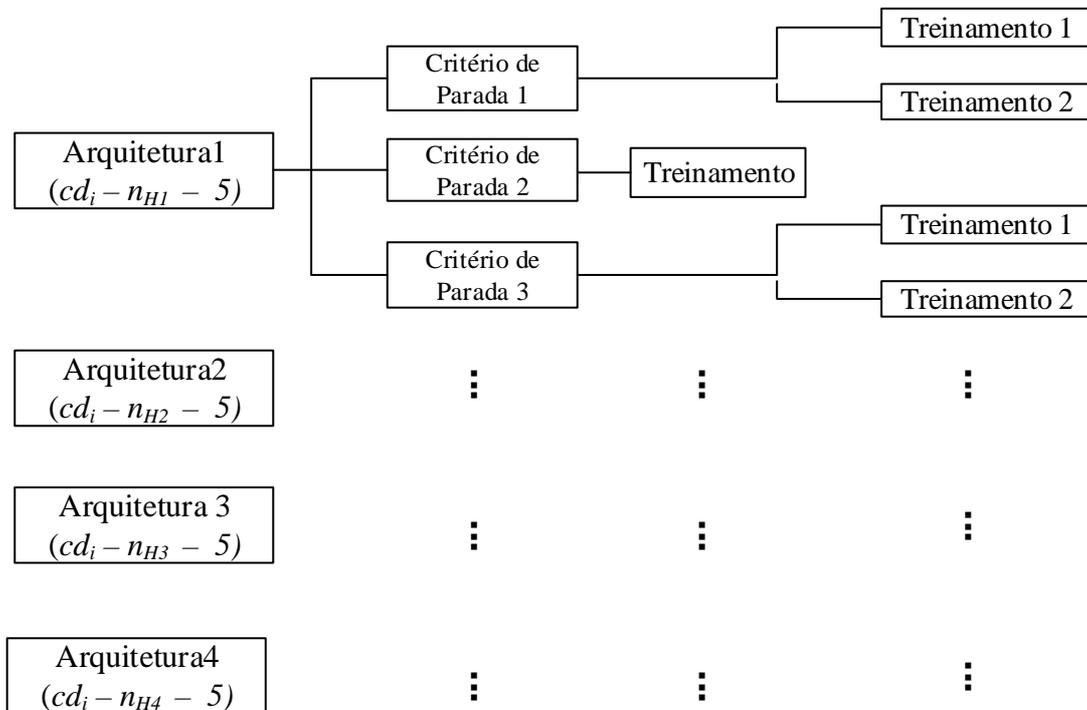


Figura 16 – Diagrama em blocos genérico aplicado em cada fase da pesquisa. cd_i representa o número de neurônios na entrada (igual ao número de características, podendo valer 10, 15 ou 22) e n_{hi} representa o número de neurônios na camada intermediária.

Cada arquitetura é treinada com os critérios de parada Erro Médio Quadrático, Parada Antecipada e Regularização, representados na Figura 16 como Critério de Parada 1, Critério de Parada 2 e Critério de Parada 3, respectivamente. O primeiro e último critério mencionados são divididos e testados com as metodologias de treinamento *leave-one-out* e *10-fold cross validation*, representadas na Figura 16 como Treinamento 1 e Treinamento 2, respectivamente. Cada etapa será detalhada neste capítulo.

A tabela 9 apresenta um sumário dos passos metodológicos implementados para cada uma das quatro arquiteturas de redes neurais testadas. Lembrando que as mesmas diferem unicamente

pelo número de neurônios da camada intermediária: a primeira arquitetura com 1 neurônio na camada intermediária, a segunda com 10, a terceira com 14 e a quarta com 45 neurônios, respectivamente.

Tabela 9 – Sumário dos passos metodológicos a serem realizados para cada uma das 4 arquiteturas de redes neurais utilizadas

Critério de parada da Rede Neural	Metodologias de treinamento e teste	Método de extração de características/número de características
Erro Médio Quadrático	<i>Leave-one-out</i>	--/conjunto de dados completo
		PCA/ 15 componentes
		PCA /10 componentes
		LDA/ 15 componentes
		LDA/ 10 componentes
	Validação cruzada (<i>10-fold</i>)	--/conjunto de dados completo
		PCA/ 15 componentes
		PCA /10 componentes
		LDA/ 15 componentes
		LDA/ 10 componentes
Parada Antecipada	Treinamento: 1/3; Validação:1/3; Teste:1/3	--/conjunto de dados completo
		PCA/ 15 componentes
		PCA /10 componentes
		LDA/ 15 componentes
		LDA/ 10 componentes
Erro Regularizado	<i>Leave-one-out</i>	--/conjunto de dados completo
		PCA/ 15 componentes
		PCA /10 componentes
		LDA/ 15 componentes
		LDA/ 10 componentes
	Validação cruzada (<i>10-fold</i>)	--/conjunto de dados completo
		PCA/ 15 componentes
		PCA /10 componentes
		LDA/ 15 componentes
		LDA/ 10 componentes

Para cada uma dessas combinações serão mostrados os seguintes resultados:

- Tabela mostrando o resultado previsto pelo médico e pelo método, para cada um dos trinta pacientes;
- Tabela de confusão dos resultados;
- Tabela-resumo contendo informações de todos os métodos.

4.2.1 O algoritmo de treinamento: Levenberg Marquardt

Nesta seção apresentam-se o detalhamento dos métodos de treinamento (Levenberg-Marquardt), os critérios de parada de cada motor, a divisão dos conjuntos do banco de dados e a arquitetura da rede.

A fase de aprendizagem é uma das fases mais críticas na construção de uma aplicação de uma rede neural bem sucedida. Segundo Lera e Pinzolas (2002), desde o desenvolvimento do método do *backpropagation*, sempre houve forte interesse na busca de novos e melhores algoritmos de treinamento.

Este algoritmo de treinamento foi apresentado por Marquardt (1963) para a estimação de parâmetros não lineares e é utilizado devido a sua ótima velocidade de treinamento quando satisfeitas suas restrições de uso. O treinamento foi feito através de uma aproximação do Hessiano (\mathbf{H}) de seu conjunto de dados pelo produto de Jacobianos (\mathbf{J}) deste mesmo conjunto:

$$\mathbf{H} = \mathbf{J}^T \mathbf{J} \quad (70)$$

O gradiente pode ser calculado como:

$$\mathbf{g} = \mathbf{J}^T \mathbf{e} \quad (71)$$

Em que \mathbf{J} é a matriz Jacobiana que contém primeiras derivadas dos erros da rede em relação aos pesos, e \mathbf{e} é um vetor de erros de rede.

A atualização dos pesos durante o treinamento da rede pode ser feita através da equação (72), bastante similar ao método de Newton para estimação de parâmetros (Lera e Pinzolas, 2002). Nesta equação, tem-se que \mathbf{x}_{k+1} é o elemento na próxima iteração, \mathbf{x}_k é o atual valor do elemento, η é a taxa de treinamento da rede e \mathbf{e} , o erro quadrático médio do problema:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - [\mathbf{J}^T \mathbf{J} + \mu \mathbf{I}]^{-1} \mathbf{J}^T \mathbf{e} \quad (72)$$

O gradiente é obtido em (72) a partir do termo $\mathbf{J}^T \mathbf{e}$. O termo $\mu \mathbf{I}$ foi acrescentado em (72) para assegurar que a matriz $(\mathbf{J}^T \mathbf{J} + \mu \mathbf{I})$ seja positiva e definida para qualquer dimensão dos vetores de entrada.

Como o método de Newton converge otimamente quando o erro é pequeno, busca-se diminuir μ com a evolução do treinamento para aproximar-se da solução desejada.

A principal restrição ao uso deste algoritmo é a utilização intensiva de memória do computador durante a fase de treinamento, o que limita sua aplicação a redes pequenas, normalmente apresentando algumas centenas de pesos a serem ajustados (Moré, 1978).

4.2.2 Critérios de parada da RNA

Conforme já mencionado, serão utilizadas três técnicas de treinamento para as redes neurais: Erro Médio Quadrático, Parada Antecipada e Regularização, representados como Critério de Parada 1, Critério de Parada 2 e Critério de Parada 3, respectivamente. Cada uma possui particularidades em relação aos seus critérios de parada (Figura 17).

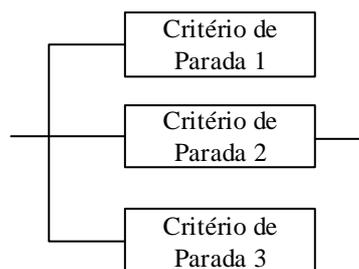


Figura 17 - Representação dos critérios de parada extraídos da Figura 16

O treinamento utilizando o Erro Médio Quadrático para quando qualquer uma dessas condições ocorrer:

- 1) O número máximo de 1000 iterações é alcançado;
- 2) O mínimo erro médio quadrático de 10^{-6} é alcançado;
- 3) μ excede μ_{max} que é igual a 10^{10} .

Segundo Wei (1994), o Erro Médio Quadrático ou Erro Quadrático Médio (EQM) é uma forma de avaliar a diferença entre um estimador e o verdadeiro valor da quantidade estimada. Ele mede a média do quadrado do erro, com o erro sendo o montante pelo qual o estimador difere da quantidade a ser estimada.

Em modelagem estatística, EQM é usado para determinar em que medida o modelo não se ajustou aos dados. Além disso, o EQM fornece um meio de escolher o melhor estimador: um EQM mínimo frequentemente, mas nem sempre, indica variação mínima e, portanto, um bom estimador (Wei, 1994).

Ter um Erro Quadrático Médio de zero (0) é o ideal, mas na maioria das situações nunca é possível. O EQM de zero significa que o estimador prevê observações com precisão perfeita.

O erro quadrático médio pode ser expresso pela diferença em distância entre os pontos de controle e os transformados a partir destes. O EQM é dado pela fórmula:

$$x_{EQM} = \sqrt{(x_t - x_m)^2 + (y_t - y_m)^2} \quad (73)$$

Onde x_t, y_t são as coordenadas transformadas e x_m, y_m são os pontos de controle.

O segundo treinamento utiliza é a Parada Antecipada (*Early Stopping*). Este treinamento para quando qualquer uma dessas condições ocorrer:

- 1) O número máximo de 1000 iterações é alcançado.
- 2) O erro médio quadrático de 10^{-6} é alcançado.

3) O erro do conjunto de Validação aumentou mais de 6 vezes desde a última vez que diminuiu.

O processo de treinamento com o auxílio da metodologia *Early Stopping* é detalhado no Capítulo 3.

O terceiro e último treinamento corresponde ao da Regularização. O critério de parada é o mesmo usado no treinamento da rede com o Erro Médio Quadrático. Ele para quando qualquer uma dessas condições ocorrer:

- 1) O número máximo de 1000 iterações é alcançado;
- 2) O mínimo erro médio quadrático de 10^{-6} é alcançado;
- 3) μ excede μ_{\max} que é igual a 10^{10} .

O processo de treinamento da rede neural utilizando o método da Regularização é detalhado no Capítulo 3.

4.2.3 Divisão do banco de dados para treinamento e teste

Detalhados os critérios de parada dos três motores utilizados na rede neural, o próximo passo é a divisão do banco de dados em conjuntos, definindo quais os dados serão utilizados para o treinamento e quais para os testes e validação.

Como a base de dados é pequena, para as Redes Neurais que utilizam o Erro Médio Quadrático e a Regularização, serão utilizadas as seguintes metodologias de treinamento e teste: *leave-one-out* e *10-fold-cross-validation* (Blum, Kalai e Langford, 1999), representadas por Treinamento 1 e Treinamento 2, respectivamente.

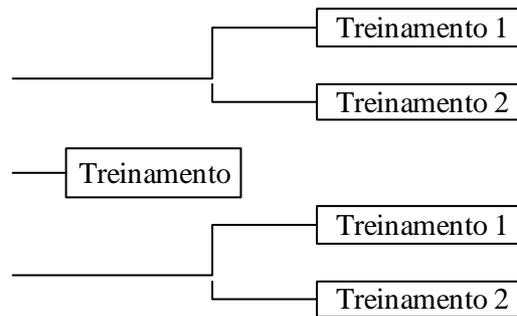


Figura 18 - Metodologias de treinamento e teste extraído da Figura 16

Essas duas metodologias são formas de validação cruzada. Na validação cruzada, assume-se que uma determinada amostra não tenha sido coletada, ou seja, elimina-se seu valor e efetua-se o treinamento a partir dos dados circundantes, podendo-se assim calcular os erros de estimativas para os pontos que são eliminados em cada etapa da validação. Após o treinamento, o valor real dessa amostra é reintroduzido no sistema e se repete o processo até que todos os valores amostrais sejam estimados dessa forma (Kohavi, 1995).

O método de validação cruzada denominado *k-fold* consiste em dividir o conjunto total de dados em k subconjuntos mutuamente exclusivos do mesmo tamanho e, a partir disto, um subconjunto é utilizado para teste e os $k-1$ restantes são utilizados para estimação dos parâmetros e calcula-se a acurácia do modelo. Este processo é realizado k vezes alternando de forma circular o subconjunto de teste (Kohavi, 1995). Neste trabalho, k assumirá o valor 10, por isso, denomina-se *10-fold cross validation*.

Na metodologia *10-fold-cross-validation*, o conjunto de 30 pacientes é dividido em 10 grupos (G1, G2, G3, G4, G5, G6, G7, G8, G9 e G10) de 3 pacientes. A rede é treinada com os dados de pacientes de 9 desses grupos e testado com os dados de pacientes do décimo grupo. Como temos 10 grupos, esse processo é repetido 10 vezes e uma média do desempenho é calculada (Figura 19). A tabela 10, em associação com a Figura 19 detalha como foram realizados a divisão dos grupos nos dez experimentos, segundo a técnica *10-fold-cross-validation*.

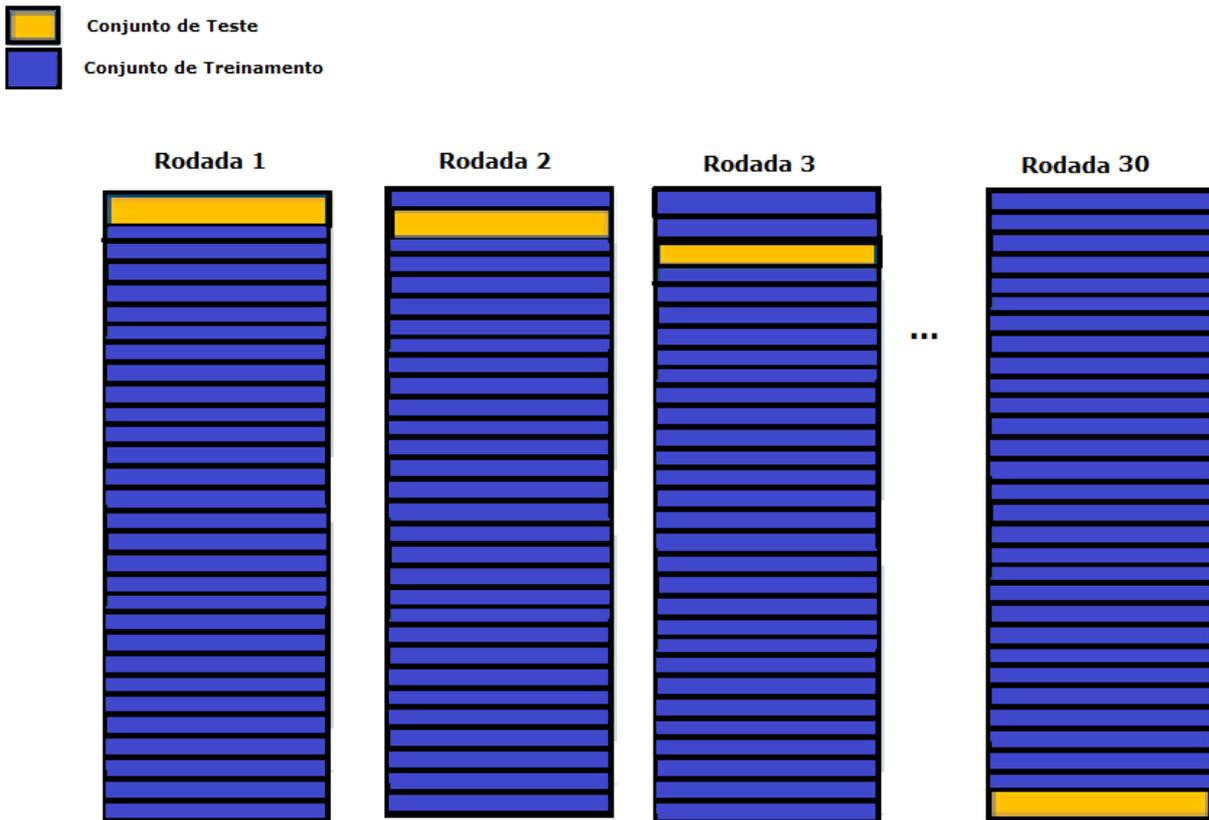


Acurácia Final = Média (Acurácia Rodada 1, Acurácia Rodada 2, ... , Acurácia Rodada 10)

Figura 19 - 10-fold Cross Validation

O segundo método implementado foi o *leave-one-out*. Este é um caso específico do *k-fold*, com k igual ao número total de dados N . Neste caso, $N=k=30$. Nesta abordagem são realizados N cálculos de erro, um para cada dado. Apesar de apresentar uma investigação completa sobre a variação do modelo em relação aos dados utilizados, este método possui um alto custo computacional, sendo indicado para situações onde poucos dados estão disponíveis (Kohavi, 1995).

Na metodologia *leave-one-out* a rede neural foi treinada com os dados de 29 pacientes e testada com os dados de 1 paciente. Como temos 30 pacientes, esse processo é repetido 30 vezes e a média do desempenho da rede é calculada, como pode ser visto na Figura 20. Ao contrário do *10-fold* em que cada grupo tinha 3 pacientes, neste caso os grupos são constituídos por 1 paciente.



Acurácia Final = Média (Acurácia Rodada 1, Acurácia Rodada 2, ... , Acurácia Rodada 30)

Figura 20 - *Leave-one-out*

No treinamento da Parada Antecipada, o conjunto de dados é separado igualmente em três conjuntos: treinamento, validação e teste. Essa separação deve ser feita de forma que cada conjunto tenha uma boa representação do conjunto de dados utilizado. Esse processo de divisão é representado na Figura 18 como “Treinamento” e está detalhado no tópico 3.2.1 do capítulo anterior.

4.2.4 Arquitetura da rede

A arquitetura da rede corresponde a uma de três camadas. O número de nós na camada de entrada é igual ao número de atributos de entrada ($cd_i = 10, 15$ ou 22) e o número de nós da camada de saída é igual ao número de classes (5) (Figura 21). Há apenas uma camada escondida que será

diferenciada em função do número de neurônios. Algumas fórmulas heurísticas sugerem valores para o número de neurônios na camada escondida da rede MLP. Foram utilizadas três regras heurísticas para chegarmos aos valores de 10, 14 e 45 nós na camada escondida. Uma RNA com 1 neurônio na camada escondida também foi utilizada ($n_{H1} = 1$), resultando assim em 4 arquiteturas de RNA. As regras heurísticas utilizadas são apresentadas a seguir.

Regra da Raiz Quadrada (Kovács, 2002): o número de neurônios da camada escondida é igual a média geométrica do número de entradas (E) e do número de saídas (S) da rede

$$n_{H2} = \sqrt{E * S} \cong 10 \quad (74)$$

Regra do Valor Médio: o número de neurônios da camada escondida é a média aritmética do número de entradas (E) e do número de saídas (S) da rede

$$n_{H3} = \frac{E + S}{2} \cong 14 \quad (75)$$

Regra de Kolmogorov (Kolmogorov, 1956): o número de neurônios da camada escondida é igual a duas vezes o número de entradas (E) somado de um.

$$n_{H4} = 2E + 1 = 45 \quad (76)$$

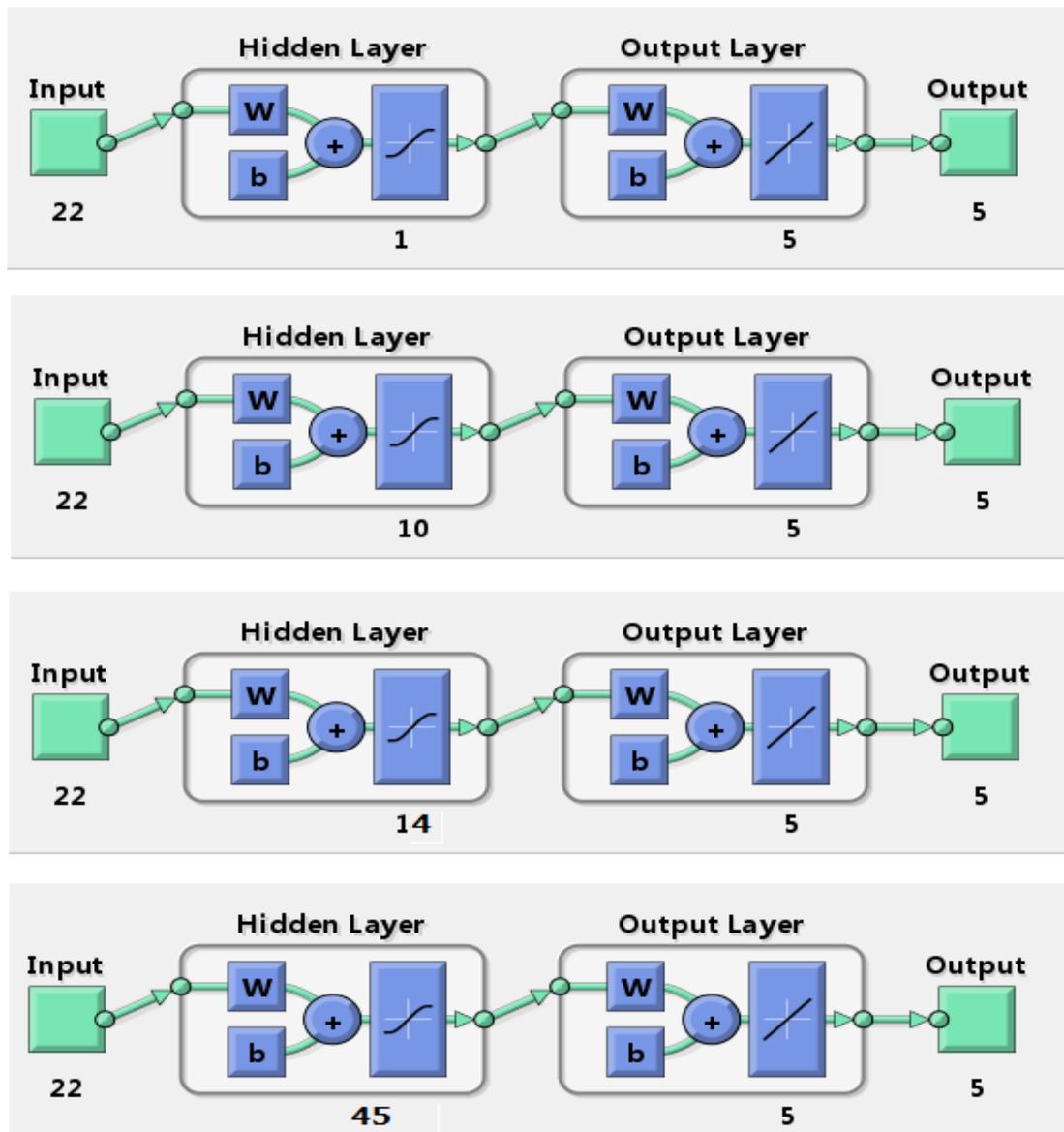


Figura 21 – As quatro arquiteturas utilizadas para o treinamento

Assim se finaliza a descrição das etapas da primeira fase. As próximas fases consistem na realização *a priori* de técnica de redução do conjunto de dados, através do uso de técnicas de redução de dimensionalidade.

Na segunda fase da pesquisa, portanto, do conjunto de dados de 30 pacientes são extraídos 10 e 15 componentes principais, respectivamente, com o auxílio da ferramenta PCA.

E por último, a terceira fase ao invés do PCA, utiliza a técnica LDA. Ambas as ferramentas foram detalhadas no capítulo 3.

4.2.5 Ambiente de Implementação

O desenvolvimento da metodologia proposta foi realizado no Laboratório de Processamento Digital de Imagens do Centro de Tecnologia Eletrônica e da Informação – CETELI da Universidade Federal do Amazonas – UFAM, utilizando-se o ambiente de desenvolvimento Matlab® (The MathWorks, Inc. Software) versão 2010 e o software Weka 3.7 desenvolvido na University of Waikato. O computador utilizado foi um PC com processador Intel Core i3 de 2.13 GHz e 4 GBytes de memória RAM.

RESULTADOS

Neste capítulo serão apresentados os resultados obtidos com o método proposto, o desempenho dos classificadores neurais e as taxas de acertos dos métodos na determinação da gravidade da malária. Inicialmente, serão mostrados os resultados da primeira fase da pesquisa, incluindo as tabelas com os resultados previstos pelo médico e pelo método para cada um dos trinta pacientes, bem como as matrizes de confusão dos resultados.

Posteriormente, são mostrados os resultados das fases utilizando o PCA e o LDA, com as comparações do diagnóstico médico com os métodos e as matrizes de confusão dos principais resultados. Por fim, será apresentada uma Tabela-resumo contendo informações percentuais de todos os métodos.

Como já mencionado, serão utilizadas matrizes de confusão para ilustrar os resultados obtidos pelos classificadores. A matriz de confusão de uma hipótese oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos (Hay, 1988).

Esta tabela valida o aprendizado supervisionado, comparando sua base de testes com a base treinada, e indica a quantidade de acertos e de erros. Este resultado é o que apresenta a acurácia do seu experimento (classificados corretamente dividido pelo total de exemplos).

O número de acertos, para cada classe, se localiza na diagonal principal da matriz. Já os demais elementos, para $i \neq j$, representam erros na classificação. A matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero uma vez que ele não comete erros.

5.1 PRIMEIRA FASE

A seguir são mostradas as tabelas de confusão dos resultados obtidos com o banco de dados completo e a taxa de acerto do método. Posteriormente, será mostrada a tabela de comparação dos resultados previstos pelo médico e pelos métodos para cada paciente.

3.2.1 Redes Neurais com Erro Médio Quadrático

- **Leave-one-out – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	12	1	0
4	0	0	2	11	0
5	0	0	0	2	0

Corretamente Classificados: 23 pacientes -

76.67%

Incorretamente Classificados: 7 pacientes -

23.33%

- **Leave-one-out – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	11	2	0
4	0	0	4	9	0
5	0	0	0	2	0

Corretamente Classificados: 20 pacientes -

66.67%

Incorretamente Classificados: 10 pacientes -

33.33%

- **Leave-one-out – Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 10 neurônios.

- **Leave-one-out – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	11	2	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 19 pacientes -

63.33%

Incorretamente Classificados: 11 pacientes -

36.67%

- **10-fold cross-validation – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	2	0	0
3	0	0	9	4	0
4	0	1	5	7	0
5	0	0	0	2	0

Corretamente Classificados: 16 pacientes -

53.33%

Incorretamente Classificados: 14 pacientes -

46.67%

- **10-fold cross-validation – Camada escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	10	3	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

- **10-fold cross-validation – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	11	2	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 19 pacientes -

63.33%

Incorretamente Classificados: 11 pacientes -

36.67%

- **10-fold cross-validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	10	3	0
4	0	0	3	8	2
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

5.1.2. Redes Neurais com Parada Antecipada

Neste caso a divisão do conjunto de dados foi feita da seguinte forma: 10 dados no conjunto de treinamento, 10 dados no conjunto de validação e 10 dados no conjunto de teste. A divisão foi feita de maneira aleatória, mas se certificando que todos os conjuntos tenham o máximo de dados representativos de todas as classes.

- **Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	5	0	0
4	0	0	4	1	0
5	0	0	0	0	0

Corretamente Classificados: 6 pacientes -

60%

Incorretamente Classificados: 4 pacientes -

40%

- **Camada Escondida: 10 neurônios**

Mesmo resultado apresentado com 1 neurônio.

- **Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 1 neurônio.

- **Camada Escondida: 45 neurônios**

Mesmo resultado apresentado com 1 neurônio.

5.1.3. Redes Neurais com Regularização

- **Leave-one-out – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	9	4	0
4	0	0	2	11	0
5	0	0	0	2	0

Corretamente Classificados: 21 pacientes -

70%

Incorretamente Classificados: 9 pacientes -

30%

- **Leave-one-out – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	12	1	0
4	0	0	0	12	1
5	0	0	0	2	0

Corretamente Classificados: 25 pacientes -

83,33%

Incorretamente Classificados: 5 pacientes -

16,67%

- **Leave-one-out – Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 10 neurônios.

- **Leave-one-out – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	8	5	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes -

56,67%

Incorretamente Classificados: 13 pacientes -

43,33%

- **10-fold cross validation – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	5	7	1
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

- **10-fold cross validation – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	5	6	2
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes -

57%

Incorretamente Classificados: 13 pacientes -

43%

- **10-fold cross validation – Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 10 neurônios.

- **10-fold cross validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	4	7	2
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

Os resultados de acurácia obtidos na primeira fase desse trabalho são mostrados na Tabela 11. Os valores de maior acurácia na predição de cada método estão em negrito e a maior acurácia entre todas as metodologias está destacada em cinza.

Tabela 11 - Acurácia dos métodos propostos na primeira fase

Critério de Parada	Metodologia de Treinamento e Teste	Neurônios na camada Intermediária	Acurácia (%)
Erro médio quadrático		1	76.7
	<i>Leave-one-out</i>	10	66.7
		14	66.7
		45	63.3
	<i>10-fold cross validation</i>	1	53.3
		10	60.0
		14	63.3
		45	60.0
Parada antecipada	10 - 10 - 10	1	
		10	
		14	60.0
		45	
Regularização		1	70.0
	<i>Leave-one-out</i>	10	83.3
		14	83.3
		45	56.7
	<i>10-fold cross validation</i>	1	60.0
		10	57.0
		14	57.0
		45	60.0

5.2 SEGUNDA FASE (PCA - 15 características)

A seguir são mostradas as tabelas de confusão dos resultados obtidos com o banco de dados reduzido por PCA para 15 características e a taxa de acerto do método.

5.2.1 Redes Neurais com Erro Médio Quadrático

- **Leave-one-out – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	2	0
3	0	0	8	5	0
4	0	0	3	10	0
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes

60%

Incorretamente Classificados: 12 pacientes

40%

- **Leave-one-out – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	0	1	0
3	0	0	8	5	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes -

56.67%

Incorretamente Classificados: 13 pacientes -

43.33%

- **Leave-one-out – Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 10 neurônios.

▪ ***Leave-one-out* – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 19 pacientes –
63.33%

Incorretamente Classificados: 11 pacientes -
36.67%

▪ ***10-fold cross-validation* – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	7	6	0
4	0	0	6	7	0
5	0	0	0	2	0

Corretamente Classificados: 14 pacientes –
46.67%

Incorretamente Classificados: 16 pacientes –
53.33%

▪ ***10-fold cross-validation* – Camada escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	7	6	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

- **10-fold cross-validation – Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 10 neurônios.

- **10-fold cross-validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	7	6	0
4	0	0	5	7	1
5	0	0	0	2	0

Corretamente Classificados: 15 pacientes -

50%

Incorretamente Classificados: 15 pacientes -

50%

5.2.2 Redes Neurais com Parada Antecipada

Neste caso a divisão do conjunto de dados foi feita da seguinte forma: 10 dados no conjunto de treinamento, 10 dados no conjunto de validação e 10 dados no conjunto de teste. A divisão foi feita de maneira aleatória, mas se certificando que todos os conjuntos tenham o máximo de dados representativos de todas as classes.

- **Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	5	0	0
4	0	0	5	0	0
5	0	0	0	0	0

Corretamente Classificados: 5 pacientes -

50%

Incorretamente Classificados: 5 pacientes -

50%

- **Camada Escondida: 10 neurônios, 14 neurônios e 45 neurônios**

Mesmo resultado apresentado com 1 neurônio.

5.2.3 Redes Neurais com Regularização

- ***Leave-one-out* – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	0	1	0
3	0	0	6	7	0
4	0	0	7	6	0
5	0	0	0	2	0

Corretamente Classificados: 13 pacientes -

40%

Incorretamente Classificados: 9 pacientes -

60%

- ***Leave-one-out* – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	7	6	0
4	0	0	4	9	0
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –

56.67%

Incorretamente Classificados: 13 pacientes –

43.33%

- ***Leave-one-out* – Camada Escondida: 14 neurônios**

Mesmo resultado apresentado com 10 neurônios.

- **Leave-one-out – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	1	1	0	0	0
3	0	0	7	6	0
4	0	0	4	9	0
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes -

56,67%

Incorretamente Classificados: 13 pacientes -

43,33%

- **10-fold cross validation – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	6	7	0
5	0	0	2	0	0

Corretamente Classificados: 17 pacientes -

56.67%

Incorretamente Classificados: 13 pacientes -

43.33%

- **10-fold cross validation – Camada Escondida: 10 neurônios**

Mesmo resultado apresentado sem PCA

- **10-fold cross validation – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	6	7	0
4	0	0	3	8	2
5	0	0	1	1	0

Corretamente Classificados: 16 pacientes -

53.33%

Incorretamente Classificados: 14 pacientes -

46.67%

▪ **10-fold cross validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	5	6	2
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

Os resultados de acurácia obtidos na segunda fase desse trabalho, utilizando PCA, com 15 características são mostrados na Tabela 12. Os valores de maior acurácia na predição de cada método estão em negrito e a maior acurácia entre todas as metodologias está destacada em cinza.

Tabela 12 – Resumo dos resultados obtidos na segunda fase com 15 características

PCA - 15 características					
Critério de Parada	Treinamento	N° de Neurônios	Acurácia		
Erro Médio Quadrático	Leave-one-out	1	60.0%		
		10	56.7%		
		14	56.7%		
		45	63.3%		
		1	46.7%		
		10	56.7%		
		14	56.7%		
		45	50.0%		
		Parada Antecipada	Conjunto de Validação	1	
				10	50.0%
14					
45					
1	40.0%				
Regularização	Leave-one-out	10	57.0%		
		14	57.0%		
		45	56.7%		
		1	57.0%		
		10	57.0%		
		14	53.0%		
		45	57.0%		

5.3 SEGUNDA FASE (PCA – 10 características)

5.3.1 Redes Neurais com Erro Médio Quadrático

- **Leave-one-out – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	0	2	0
3	0	0	10	3	0
4	0	0	3	10	0
5	0	0	0	2	0

Corretamente Classificados: 20 pacientes –

66.67%

Incorretamente Classificados: 10 pacientes -

33.33%

- **Leave-one-out – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	11	2	0
4	0	0	4	7	2
5	0	0	0	2	0

Corretamente Classificados: 20 pacientes –

66.67%

Incorretamente Classificados: 10 pacientes –

33.33%

- **Leave-one-out – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	11	2	0
4	0	0	3	8	0
5	0	0	0	2	0

Corretamente Classificados: 21 pacientes –

70%

Incorretamente Classificados: 9 pacientes –

30%

- **Leave-one-out – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	11	2	0
4	0	0	5	7	1
5	0	0	0	2	0

Corretamente Classificados: 20 pacientes -

66.67%

Incorretamente Classificados: 10 pacientes -

33.33%

- **10-fold cross-validation – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	2	0	0
3	0	0	7	6	0
4	0	0	5	8	0
5	0	0	0	2	0

Corretamente Classificados: 15 pacientes -

50%

Incorretamente Classificados: 15 pacientes -

50%

- **10-fold cross-validation – Camada escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	7	6	0
4	0	0	4	7	2
5	0	0	0	2	0

Corretamente Classificados: 16 pacientes -

53.33%

Incorretamente Classificados: 14 pacientes -

46.67%

- **10-fold cross-validation – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	0	1	0
3	0	0	8	5	0
4	0	0	3	8	2
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

- **10-fold cross-validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	4	7	2
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

5.3.2 Redes Neurais com Parada Antecipada

Mesmo resultado apresentado com 15 características.

5.3.3 Redes Neurais com Regularização

- **Leave-one-out – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	6	7	0
5	0	0	2	0	0

Corretamente Classificados: 18 pacientes –
60%

Incorretamente Classificados: 12 pacientes –
40%

▪ **Leave-one-out – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	3	9	1
5	0	0	0	2	0

Corretamente Classificados: 19 pacientes –
63.33%

Incorretamente Classificados: 11 pacientes –
36.67%

▪ **Leave-one-out – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	4	8	1
5	0	0	1	1	0

Corretamente Classificados: 19 pacientes –
63.33%

Incorretamente Classificados: 11 pacientes –
36.67%

▪ **Leave-one-out – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	4	6	3
5	0	0	1	1	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

- **10-fold cross validation – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	5	7	1
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

- **10-fold cross validation – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	4	8	1
5	0	0	2	0	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

- **10-fold cross validation – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	3	10	0
5	0	0	2	0	0

Corretamente Classificados: 18 pacientes –

66.67%

Incorretamente Classificados: 12 pacientes –

33.33%

- **10-fold cross validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	5	6	2
5	0	0	1	1	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

Os resultados de acurácia obtidos na segunda fase, utilizando PCA, com 10 características são mostrados na Tabela 13. Os valores de maior acurácia na predição de cada método estão em negrito e a maior acurácia entre todas as metodologias está destacada em cinza.

Tabela 13 - Resumo dos resultados obtidos na segunda fase com 10 características

PCA - 10 características					
Critério de Parada	Treinamento	N° de Neurônios	Acurácia		
Erro Médio Quadrático		1	66.7%		
		Leave-one-out	10	66.7%	
			14	70.0%	
		45	66.7%		
		Parada Antecipada	Conjunto de Validação	1	50.0%
				10	50.0%
14	50.0%				
45	50.0%				
Regularização		1	60.0%		
		Leave-one-out	10	63.3%	
			14	63.3%	
		45	56.7%		
		10-fold Cross Validation		1	60.0%
				10	60.0%
14	66.7%				
45	56.7%				

5.4 TERCEIRA FASE (LDA - 15 características)

A seguir são mostradas as tabelas de confusão dos resultados obtidos com o banco de dados reduzido por LDA para 15 características e a taxa de acerto do método.

5.4.1 Redes Neurais com Erro Médio Quadrático

- **Leave-one-out – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	3	10	0
5	0	0	0	2	0

Corretamente Classificados: 20 pacientes

66.67%

Incorretamente Classificados: 10 pacientes

33.33%

- **Leave-one-out – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -

60%

Incorretamente Classificados: 12 pacientes -

40%

▪ ***Leave-one-out* – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	0	1	0
3	0	0	8	5	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –

56.67%

Incorretamente Classificados: 13 pacientes –

43.33%

▪ ***Leave-one-out* – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	4	8	1
5	0	0	0	1	1

Corretamente Classificados: 20 pacientes –

66.67%

Incorretamente Classificados: 10 pacientes –

33.3%

▪ ***10-fold cross-validation* – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	7	6	0
4	0	0	6	7	0
5	0	0	0	2	0

Corretamente Classificados: 14 pacientes –

46.67%

Incorretamente Classificados: 16 pacientes –

53.33%

▪ **10-fold cross-validation – Camada escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	7	6	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

▪ **10-fold cross-validation – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	7	6	0
4	0	0	4	8	1
5	0	0	0	1	1

Corretamente Classificados: 18 pacientes –
60%

Incorretamente Classificados: 12 pacientes –
40%

▪ **10-fold cross-validation – Camada Escondida: 45 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	7	6	0
4	0	0	5	7	1
5	0	0	0	2	0

Corretamente Classificados: 15 pacientes -
50%

Incorretamente Classificados: 15 pacientes -
50%

5.4.2 Redes Neurais com Parada Antecipada

Mesmo resultado apresentado com PCA.

5.4.3 Redes Neurais com Regularização

- ***Leave-one-out* – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	7	6	0
4	0	0	4	8	1
5	0	0	0	2	0

Corretamente Classificados: 17 pacientes –
56.67%

Incorretamente Classificados: 13 pacientes –
43.33%

- ***Leave-one-out* – Camada Escondida: 10 neurônios, 14 neurônios e 45 neurônios**

Mesmo resultado apresentado com PCA – 15 características.

- ***10-fold cross validation* – Camada Escondida: 1 neurônio**

Mesmo resultado apresentado com o banco de dados completo e sem redução de características.

- ***10-fold cross validation* – Camada Escondida: 10 neurônios, 14 neurônios, 45 neurônios**

Mesmo resultado apresentado com PCA – 15 características.

Os resultados de acurácia obtidos na terceira fase, utilizando LDA, com 15 características são mostrados na Tabela 14. Os valores de maior acurácia na predição de cada método estão em negrito e a maior acurácia entre todas as metodologias está destacada em cinza.

Tabela 14 - Resumo dos resultados obtidos na terceira fase utilizando LDA com 15 características

LDA - 15 características			
Critério de Parada	Treinamento	N° de Neurônios	Acurácia
Erro Médio Quadrático		1	66.7%
	Leave-one-out	10	60.0%
		14	56.7%
		45	66.7%
	10-fold Cross Validation	1	46.7%
10		56.7%	
14		60.0%	
	45	50.0%	
Parada Antecipada		1	
	Conjunto de Validação	10	50.0%
		14	
		45	
Regularização		1	56.7%
	Leave-one-out	10	56.7%
		14	56.7%
		45	56.7%
	10-fold Cross Validation	1	60.0%
		10	56.7%
14		53.0%	
	45	56.7%	

5.5 TERCEIRA FASE (LDA – 10 características)

5.5.1 Redes Neurais com Erro Médio Quadrático

- ***Leave-one-out* – Camada Escondida: 1 neurônio**

Mesmo resultado apresentado com LDA – 15 características.

- ***Leave-one-out* – Camada Escondida: 10 neurônios e 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	11	2	0
4	0	0	3	8	0
5	0	0	0	2	0

Corretamente Classificados: 21 pacientes –
70%

Incorretamente Classificados: 9 pacientes –
30%

- ***Leave-one-out* – Camada Escondida: 45 neurônios**

Mesmo resultado apresentado com LDA – 15 características.

- ***10-fold cross-validation* – Camada Escondida: 1 neurônio**

Mesmo resultado apresentado com o banco de dados completo, sem extração de características.

- ***10-fold cross-validation* – Camada escondida: 10 neurônios, 14 neurônios e 45 neurônios**

Mesmo resultado apresentado com PCA – 10 características.

5.5.2 Redes Neurais com Parada Antecipada

Mesmo resultado apresentado com 15 características.

5.5.3 Redes Neurais com Regularização

- ***Leave-one-out* – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	6	7	0
5	0	0	2	0	0

Corretamente Classificados: 18 pacientes –

60%

Incorretamente Classificados: 12 pacientes –

40%

- ***Leave-one-out* – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	3	9	1
5	0	0	0	2	0

Corretamente Classificados: 19 pacientes –

63.33%

Incorretamente Classificados: 11 pacientes –

36.67%

- ***Leave-one-out* – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	11	2	0
4	0	0	3	8	0
5	0	0	0	2	0

Corretamente Classificados: 21 pacientes –

70%

Incorretamente Classificados: 9 pacientes –

30%

- ***Leave-one-out* – Camada Escondida: 45 neurônios**

Mesmo resultado apresentado em LDA – 15 características.

- ***10-fold cross validation* – Camada Escondida: 1 neurônio**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	9	4	0
4	0	0	5	7	1
5	0	0	0	2	0

Corretamente Classificados: 18 pacientes -
60%

Incorretamente Classificados: 12 pacientes -
40%

- ***10-fold cross validation* – Camada Escondida: 10 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	4	8	1
5	0	0	2	0	0

Corretamente Classificados: 18 pacientes -
60%

Incorretamente Classificados: 12 pacientes -
40%

- ***10-fold cross validation* – Camada Escondida: 14 neurônios**

Classes	Classificado como				
	1	2	3	4	5
1	0	0	0	0	0
2	0	2	0	0	0
3	0	0	8	5	0
4	0	0	3	10	0
5	0	0	2	0	0

Corretamente Classificados: 18 pacientes –
66.67%

Incorretamente Classificados: 12 pacientes –
33.33%

- **10-fold cross validation – Camada Escondida: 45 neurônios**

Mesmo resultado apresentado com o banco de dados completo, sem extração de características.

Os resultados de acurácia obtidos na terceira fase, utilizando LDA, com 10 características são mostrados na Tabela 15. Os valores de maior acurácia na predição de cada método estão em negrito e a maior acurácia entre todas as metodologias está destacada em cinza.

Tabela 15 - Resumo dos resultados obtidos na terceira fase utilizando LDA com 10 características

LDA - 10 características					
Critério de Parada	Treinamento	N° de Neurônios	Acurácia		
Erro Médio Quadrático	Leave-one-out	1	66.7%		
		10	70.0%		
		14	70.0%		
		45	66.7%	
	10-fold Cross Validation		1	53.3%
			10	53.3%
.....		14	56.7%		
.....		45	56.7%		
Parada Antecipada	Conjunto de Validação	1			
		10	50.0%		
		14			
.....		45			
Regularização	Leave-one-out	1	60.0%		
		10	63.0%		
		14	70.0%		
		45	56.7%	
	10-fold Cross Validation		1	60.0%
			10	60.0%
.....		14	67.0%		
.....		45	60.0%		

DISCUSSÃO E CONCLUSÕES

Importantes critérios que definem uma estrutura de uma RNA, como a quantidade de entradas, camadas intermediárias e parâmetros de treinamento, são importantes pré-requisitos para uma boa avaliação e ajuste do modelo.

A escolha da arquitetura da RNA foi uma rede neural multicamadas devido a sua grande aplicabilidade em problemas como este que envolvem aproximações. A maior dificuldade da RNA foi obter os parâmetros que melhor se comportam com os dados apresentados. Para treinar os dados, utilizaram-se 1, 10, 14 e 45 neurônios na camada intermediária. As taxas de aprendizagem e quantidade de ciclos foram obtidos após exaustivos testes.

Outro fator determinante foi a definição dos critérios de parada do treinamento da RNA, a Parada Antecipada, a Regularização e o Erro Médio Quadrático. Usando Parada Antecipada e Regularização, o ideal é treinar a rede a partir de várias condições iniciais diferentes. É possível que um ou outro método possa falhar em certas circunstâncias. Ao testar várias condições iniciais diferentes, busca-se otimizar o desempenho da rede.

Na Parada Antecipada, a escolha do conjunto de validação é importante. O conjunto de validação deve ser representativo de todos os pontos do conjunto de treinamento. Quando se usa regularização Bayesiana, é importante treinar a rede até que ela chegue a convergência. O erro da soma dos quadrados, os pesos da soma dos quadrados, e o número efetivo de parâmetros devem atingir valores constantes quando a rede convergir. Esse processo geralmente leva mais tempo que a Parada Antecipada.

Na Tabela 16 encontram-se os melhores resultados obtidos com cada critério de parada. A regularização, com *leave-one-out*, com 10 ou 14 neurônios na camada intermediária, apresentou a melhor acurácia entre todos os métodos de 83.3%.

Quando o conjunto de dados é pequeno e se está treinando redes de aproximação de funções, a Regularização Bayesiana oferece melhor desempenho na generalização que a Parada Antecipada. Isso ocorre porque a Regularização Bayesiana não exige que um conjunto de dados de validação seja separado do conjunto de treinamento, utilizando todos os dados. Isso justifica o desempenho da Regularização Bayesiana ter sido superior ao da Parada Antecipada.

Pelos resultados obtidos, verificou-se que a utilização de 10 ou 14 neurônios na camada intermediária se mostrou a mais acertada na maioria dos experimentos. O método *leave-one-out* também se destacou no decorrer dos experimentos, que apesar do custo computacional, tira o máximo proveito dos dados.

Tabela 16 - Melhores resultados de cada critério de parada

	Critério de Parada	Treinamento	N° de Neurônios	Acurácia
Banco de dados completo	Erro Médio Quadrático	Leave-one-out	1	76.7%
	Parada Antecipada	Conjunto de Validação	1	60.0%
			10	
			14	
			45	
Regularização	Leave-one-out	10	83.3%	
			14	83.3%
PCA - 10 características	Erro Médio Quadrático	Leave-one-out	14	70.0%
	Parada Antecipada	Conjunto de Validação	1	50.0%
			10	
			14	
			45	
Regularização	10-fold Cross Validation	14	66.7%	
PCA - 15 características	Erro Médio Quadrático	Leave-one-out	45	63.3%
	Parada Antecipada	Conjunto de Validação	1	50.0%
			10	
			14	
			45	
	Regularização	Leave-one-out	10	57.0%
			14	57.0%
10-fold Cross Validation			1	57.0%
		10	57.0%	
		45	57.0%	
LDA - 10 características	Erro Médio Quadrático	Leave-one-out	10	70.0%
			14	70.0%
	Parada Antecipada	Conjunto de Validação	1	50.0%
			10	
			14	
45				
Regularização	Leave-one-out	14	70.0%	

		Erro Médio Quadrático	Leave-one-out	45	66.7%
LDA - 15 características	Parada Antecipada			1	50.0%
				10	
				14	
				45	
		Regularização	10-fold Cross Validation	1	60.0%

Vale ressaltar que o método proposto difere fundamentalmente de outros analisados no capítulo de referências bibliográficas, principalmente do método de Uzoka, no que diz respeito à sequência das etapas utilizadas. A divisão macro do estudo foi dividida em três fases, que se diferenciavam de acordo com a maneira que o banco de dados era utilizado. A primeira fase foi utilizando o banco de dados completo com os 22 sintomas (características) de cada um dos 30 pacientes, a segunda fase, com auxílio do PCA, com 15 e 10 características. E por último, utilizando a ferramenta LDA, a RNA também foi treinada com 15 e 10 características.

Na Tabela 17 mostramos uma comparação entre o melhor resultado obtido nas três fases, com os melhores resultados obtidos por (Uzoka, Osuji e Obot, 2011). Com *Fuzzy*, eles obtiveram 80% de acurácia e com AHP, 76,6%. Nessa tabela é mostrado também o resultado da avaliação da gravidade da malária feita por um especialista.

A partir dessa tabela calcula-se a acurácia do método e avaliação da gravidade: Para cada paciente, compara-se o resultado do método com a avaliação do especialista. Se os dois valores forem positivos contabiliza-se um acerto do método.

Tabela 17 - Comparação entre o método proposto (redes neurais regularizadas) e os apresentados em (Uzoka, Osuji e Obot, 2011).

Método de Avaliação	Paciente																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Diagnóstico Médico	4	3	4	4	3	3	4	3	4	3	3	3	2	3	5	4	3	3	4	4	3	4	3	5	4	4	3	2	4	4
(Uzoka, Osuji e Obot, 2011).	4	3	3	4	3	3	3	3	4	3	3	2	2	3	4	4	3	4	4	4	3	4	3	5	3	4	3	2	4	4
Método proposto	4	3	4	4	3	3	4	3	4	3	3	3	2	3	4	4	3	3	4	4	3	5	3	4	4	4	3	3	4	5

Os diagnósticos foram submetidos a um teste de hipóteses Qui Quadrado, χ^2 . Este se destina a encontrar um valor da dispersão para duas variáveis nominais, avaliando a associação existente entre variáveis qualitativas. O princípio básico deste método é comparar proporções, isto é, as possíveis divergências entre as frequências observadas e esperadas para um certo evento.

Evidentemente, pode-se dizer que dois grupos se comportam de forma semelhante se as diferenças entre as frequências observadas e as esperadas em cada categoria forem muito pequenas, próximas a zero.

Na comparação dos resultados obtidos com os apresentados por Uzoka *et al.* (2011), utilizando um grau de liberdade 1, o valor de χ^2 igual a 0,48 não é estatisticamente significativo ao nível de 5%, ou seja, apesar do método da regularização mostrar-se melhor do que a rede *fuzzy*, a diferença estatística não foi significativa.

Em suma, todos os objetivos do trabalho foram realizados com sucesso. Além disso, foram identificadas as principais ferramentas de Suporte à Decisão Clínica em malária publicadas, caracterizamos e comparamos cada estilo de critério de parada e métodos de extração de características associados a redes neurais na avaliação da gravidade da malária.

Como trabalhos futuros, vislumbra-se a utilização de outras técnicas de reconhecimento de padrões, como máquinas de vetores de suporte. Mas principalmente, espera-se expandir o banco de dados a partir do estabelecimento de cooperação com instituições de saúde que atuam no diagnóstico e tratamento da malária no estado do Amazonas, o que não foi possível para este trabalho.

REFERÊNCIAS

- Ahmadian, L. et al. The role of standardized data and terminological systems in computerized clinical decision support systems: literature review and survey. **Int J Med Inform**, v. 80, n. 2, p. 81-93, Feb 2011. ISSN 1386-5056.
- Almeida, P. P. d. Aplicação do método AHP-Processo Analítico Hierárquico-à seleção de helicópteros para apoio logístico à exploração e produção de petróleo "offshore". 2002.
- Barros, A. M. et al. Artificial Neural Networks and Bayesian Networks as supporting tools for diagnosis of asymptomatic malaria. e-Health Networking Applications and Services (Healthcom), 2010 12th IEEE International Conference on, 2010, 1-3 July 2010. p.106-111.
- Bassanezi, R.; Barros, L. Tópicos de lógica fuzzy e biomatemática. **Coleção IMECC: textos didáticos**, v. 5, 2006.
- Berner, E. S.; La Lande, T. J. Overview of clinical decision support systems. In: (Ed.). **Clinical decision support systems**: Springer, 2007. p.3-22. ISBN 0387339140.
- Bilobrovec, M.; Marçal, R. F. M.; Kovalesski, J. L. Implementação de um sistema de controle inteligente utilizando a lógica fuzzy. **XI SIMPEP, Bauru/Brasil**, 2004.
- Bishop, C. M. Neural networks for pattern recognition. 1995.
- Blum, A.; Kalai, A.; Langford, J. Beating the hold-out: Bounds for k-fold and progressive cross-validation. Proceedings of the twelfth annual conference on Computational learning theory, 1999, ACM. p.203-208.
- Bogich, T. L. et al. Using network theory to identify the causes of disease outbreaks of unknown origin. **J R Soc Interface**, v. 10, n. 81, p. 20120904, Apr 6 2013. ISSN 1742-5662.
- Braga, A. P.; Carvalho, A. P. L.; Ludermir, T. B. **Redes Neurais Artificiais: teoria e aplicações**. Rio de Janeiro: Livros Técnicos e Científicos: 262p p. 2007.
- Camargo, E. P. Malária, maleita, paludismo. **Ciência e cultura**, v. 55, n. 1, p. 26-29, 2003. ISSN 0009-6725.
- Caudill, M.; Butler, C. **Naturally intelligent systems**. MIT press, 1992. ISBN 0262531135.
- Chandramohan, D. et al. A clinical algorithm for the diagnosis of malaria: results of an evaluation in an area of low endemicity. **Trop Med Int Health**, v. 6, n. 7, p. 505-10, Jul 2001. ISSN 1360-2276 (Print) 1360-2276.
- Chandramohan, D.; Jaffar, S.; Greenwood, B. Use of clinical algorithms for diagnosing malaria. **Trop Med Int Health**, v. 7, n. 1, p. 45-52, Jan 2002. ISSN 1360-2276 (Print) 1360-2276.

Dan Foresee, F.; Hagan, M. T. Gauss-Newton approximation to Bayesian learning. *Neural Networks*, 1997., International Conference on, 1997, IEEE. p.1930-1935.

Doan, C.; Liong, S. Generalization for multilayer neural network: Bayesian regularization or early stopping. *Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference*, 2004.

Falas, T. Neural networks in empirical accounting research: an alternative to statistical models. **Neural network world**, v. 5, n. 4, p. 419-432, 1995.

Ferreira, A. R. L. Pesquisa de anticorpos anti-Plasmodium spp em indivíduos com estadia em zona endêmica de Malária. 2013.

Ferreira, L. **Aplicação de sistemas fuzzy e neuro-fuzzy para predição da temperatura retal de frangos de corte**. 2009. Dissertação de mestrado, Universidade Federal de Lavras, Lavras, Minas Gerais, Brasil

Garg, A. X. et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. **Jama**, v. 293, n. 10, p. 1223-1238, 2005. ISSN 0098-7484.

Genton, B. et al. Malaria: how useful are clinical criteria for improving the diagnosis in a highly endemic area? **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 88, n. 5, p. 537-541, 1994. ISSN 0035-9203.

Gomes, M. et al. Symptomatic identification of malaria in the home and in the primary health care clinic. **Bulletin of the World Health Organization**, v. 72, n. 3, p. 383, 1994.

Hay, A. The derivation of global estimates from a confusion matrix. **International Journal of Remote Sensing**, v. 9, n. 8, p. 1395-1398, 1988. ISSN 0143-1161.

Haykin, S.; Network, N. A comprehensive foundation. **Neural Networks**, v. 2, n. 2004, 2004.

Hotelling, H. **Analysis of a Complex of Statistical Variables Into Principal Components**. *Journal of Educational Psychology*. 24: 417-441 p. 1933.

Iliffe, S. et al. Design and implementation of a computer decision support system for the diagnosis and management of dementia syndromes in primary care. **Methods of information in medicine**, v. 41, n. 2, p. 98-104, 2002. ISSN 0026-1270.

Jolliffe, I. T. **Principal Component Analysis** 2002.

Junior, R.; Azevedo, A. **Acessibilidade e mobilidade na estimativa de um índice de potencial de viagens utilizando redes neurais artificiais e sistemas de informações geográficas**. 2000. Universidade de São Paulo

Kasabov, N.; Hu, Y. Integrated optimisation method for personalised modelling and case studies for medical decision support. **International Journal of Functional Informatics and Personalised Medicine**, v. 3, n. 3, p. 236-256, 2010. ISSN 1756-2104.

Kim, S.-J.; Magnani, A.; Boyd, S. Robust fisher discriminant analysis. *Advances in Neural Information Processing Systems*, 2005. p.659-666.

Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Ijcai*, 1995. p.1137-1145.

Kolmogorov, A. On the representation of continuous functions of several variables by superpositions of continuous functions of a smaller number of variables. *Dokl. Akad. Nauk SSSR*, 1956. p.179-182.

Kong, G.; Xu, D.-L.; Yang, J.-B. Clinical decision support systems: a review on knowledge representation and inference under uncertainties. **International Journal of Computational Intelligence Systems**, v. 1, n. 2, p. 159-167, 2008. ISSN 1875-6891.

Koutroumbas, K. **Pattern Recognition**. 4th. Academic Press, 2008. 984 ISBN 978-1-59749-272-0.

Kovács, Z. L. **Redes neurais artificiais**. Editora Livraria da Fisica, 2002. ISBN 8588325144.

Lera, G.; Pinzolas, M. Neighborhood based Levenberg-Marquardt algorithm for neural network training. **Neural Networks, IEEE Transactions on**, v. 13, n. 5, p. 1200-1203, 2002. ISSN 1045-9227.

Likert, R. A technique for the measurement of attitudes. **Archives of psychology**, 1932.

Luxemburger, C. et al. Clinical features cannot predict a diagnosis of malaria or differentiate the infecting species in children living in an area of low transmission. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 92, n. 1, p. 45-49, 1998. ISSN 0035-9203.

MacKay, D. J. A practical Bayesian framework for backpropagation networks. **Neural computation**, v. 4, n. 3, p. 448-472, 1992.

Mehta, K. et al. Severe acute renal failure in malaria. **Journal of postgraduate medicine**, v. 47, n. 1, p. 24, 2001. ISSN 0022-3859.

Moody, J. et al. A simple weight decay can improve generalization. **Advances in neural information processing systems**, v. 4, p. 950-957, 1995.

Moré, J. J. The Levenberg-Marquardt algorithm: implementation and theory. In: (Ed.). **Numerical analysis**: Springer, 1978. p.105-116. ISBN 3540085386.

Muhe, L. et al. Clinical algorithm for malaria during low and high transmission seasons. **Archives of disease in childhood**, v. 81, n. 3, p. 216-220, 1999. ISSN 1468-2044.

Neto, J. M. M.; Moita, G. C. Uma introdução à análise exploratória de dados multivariados. **Química Nova**, v. 21, n. 4, p. 467-469, 1998.

Nguyen, D.; Widrow, B. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. *Neural Networks*, 1990., 1990 IJCNN International Joint Conference on, 1990, IEEE. p.21-26.

Olaleye, B. et al. Clinical predictors of malaria in Gambian children with fever or a history of fever. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 92, n. 3, p. 300-304, 1998. ISSN 0035-9203.

OLIVEIRA, H. d.; Amendola, M.; NääS, I. d. A. Estimativa das condições de conforto térmico para avicultura de postura usando a teoria dos conjuntos fuzzy. **Engenharia Agrícola**, v. 25, n. 2, p. 300-307, 2005.

Pandey, B.; Mishra, R. Knowledge and intelligent computing system in medicine. **Computers in biology and medicine**, v. 39, n. 3, p. 215-230, 2009. ISSN 0010-4825.

Pearson, K. **On lines and planes of closest fit to systems of points in space.** *Philosophical Magazine*. 2: 559-572 p. 1901.

Pereira, C. A. A. Priorização de investimentos em uma cadeia logística completa. 2000.

Prechelt, L. Automatic early stopping using cross validation: quantifying the criteria. **Neural Networks**, v. 11, n. 4, p. 761-767, 1998. ISSN 0893-6080.

Ramnarayan, P.; Britto, J. Paediatric clinical decision support systems. **Archives of disease in childhood**, v. 87, n. 5, p. 361-362, 2002. ISSN 1468-2044.

Redd, S. et al. Clinical algorithm for treatment of Plasmodium falciparum malaria in children. **The Lancet**, v. 347, n. 8996, p. 223-227, 1996. ISSN 0140-6736.

Regazzi, A. Análise multivariada. **Viçosa, MG: Universidade Federal de Viçosa**, v. 2, 2002.

Rooth, I.; Björkman, A. Fever episodes in a holoendemic malaria area of Tanzania: parasitological and clinical findings and diagnostic aspects related to malaria. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 86, n. 5, p. 479-482, 1992. ISSN 0035-9203.

Saaty, T. L. **What is the analytic hierarchy process?** Springer, 1988. ISBN 3642835570.

Schiassi, L. et al. Metodologia fuzzy aplicada à avaliação do aumento da temperatura corporal em frangos de corte. **Engenharia na agricultura**, v. 16, p. 181-191, 2008.

Schmidt, A. **Processo de apoio à tomada de decisão abordagens: AHP e Macbeth. 1995. 96p.** 1995. Dissertação (Mestrado em Engenharia), Universidade Federal de Santa Catarina, Florianópolis, SC

Seising, R. From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis. **Artificial Intelligence in Medicine**, v. 38, n. 3, p. 237-256, 2006. ISSN 0933-3657.

Sejvar, J. J. The evolving epidemiology of viral encephalitis. **Current opinion in neurology**, v. 19, n. 4, p. 350-357, 2006. ISSN 1350-7540.

Suan, O. L. Computer-aided diagnosis and treatment of malaria: The IMEX system. **Computers in Biology and Medicine**, v. 20, n. 5, p. 361-372, 1990. Disponível em: < [http://www.computersinbiologyandmedicine.com/article/0010-4825\(90\)90016-I/abstract](http://www.computersinbiologyandmedicine.com/article/0010-4825(90)90016-I/abstract) >. Acesso em: 2014/10/02.

Subramanian, V.; Hung, M. S.; Hu, M. Y. An experimental evaluation of neural networks for classification. **Computers & operations research**, v. 20, n. 7, p. 769-782, 1993. ISSN 0305-0548.

Toth-Pal, E. et al. A guideline-based computerised decision support system (CDSS) to influence general practitioners management of chronic heart failure. **Informatics in primary care**, v. 16, n. 1, p. 29-39, 2008. ISSN 1476-0320.

Tubb, N. A development path to success in neural computing. **Expert Systems Applications**, v. 9, n. 5, p. 5-9, 1993.

Uzoka, F.-M. E. et al. An experimental comparison of fuzzy logic and analytic hierarchy process for medical decision support systems. **Computer Methods and Programs in Biomedicine**, v. 103, n. 1, p. 10-27, 7// 2011. ISSN 0169-2607. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S016926071000146X> >.

Uzoka, F.-M. E.; Osuji, J.; Obot, O. Clinical decision support system (DSS) in the diagnosis of malaria: A case comparison of two soft computing methodologies. **Expert Systems with Applications**, Langford Lane, Kidlington, Oxford, OX5 1GB, United Kingdom, v. 38, n. 3, p. 1537-1553, 2011. ISSN 09574174. Disponível em: < <http://dx.doi.org/10.1016/j.eswa.2010.07.068> >.

Varella, C. A. A. Análise de componentes principais. **Universidade Federal Rural do Rio de Janeiro**, Disponível em: <http://www.ufrrj.br/institutos/it/deng/varella/Downloads>. Acessado em, v. 18, 2008.

Weber, M. et al. Evaluation of an algorithm for the integrated management of childhood illness in an area with seasonal malaria in the Gambia. **Bulletin of the World Health Organization**, v. 75, n. Suppl 1, p. 25, 1997.

Wei, W. W.-S. **Time series analysis**. Addison-Wesley publ, 1994. ISBN 0201159112.

Weir, C. et al. Cluster-randomized, controlled trial of computer-based decision support for selecting long-term anti-thrombotic therapy after acute ischaemic stroke. **QJM: monthly journal of the Association of Physicians**, v. 96, n. 2, p. 143-153, 2003. ISSN 1460-2725.

Zadeh, L. A. Fuzzy sets. **Information and control**, v. 8, n. 3, p. 338-353, 1965. ISSN 0019-9958.

World Health Organization – who. **Global Malaria Mapper – WHO World Malaria Report (2012)**. Disponível em:

<http://www.who.int/malaria/publications/world_malaria_report_2012/en/>. Acesso em: 22 de abril de 2014.

ROTINAS DESENVOLVIDAS NO MATLAB

A seguir estão descritas cada uma das rotinas em Matlab que foram desenvolvidos nesta dissertação.

Cada função tem uma ajuda que pode ser acessada a partir da janela de comando do Matlab usando a função *help*. O suporte de cada função é dividido em uma breve explicação da tarefa executada, seguida de sintaxe. Para qualquer outra solicitação larissaalmeida68@gmail.com.

Função	newff
Propósito	Cria uma rede <i>feed-forward backpropagation</i>
Sintaxe	<code>net=newff(ptreinamento,ttreinamento,{10},{}, 'trainlm')</code>
Descrição	<p>ptreinamento matriz com os vetores de entrada</p> <p>ttreinamento matriz com os vetores desejados de saída</p> <p>{10} número de neurônios na camada intermediária</p> <p>{ } função de transferência <i>default</i>: 'tansig' para camadas escondidas e 'purelin' para a camada de saída</p> <p>'trainlm' algoritmo de <i>backpropagation</i> Levenberg-Marquardt</p>

Além do 'trainlm', foram utilizados também:

'trainscg'	função de treinamento da rede que atualiza os valores dos pesos de acordo com o método do gradiente conjugado
'trainbr'	algoritmo de <i>backpropagation</i> com Regularização Bayesiana

Função	function_PCA
Propósito	Realizar a extração das componentes principais
Sintaxe	[autovetor] = function_PCA(pp,options)
Descrição	pp matriz com os vetores de entrada options a dimensão do espaço reduzido. Caso seja zero, as dimensões serão mantidas. O <i>default</i> é zero.

Ele retorna as seguintes variáveis:

autovetor cada coluna é multiplicada por um vetor-linha do banco de dados a fim de que se obtenha a matriz de dados reduzida.

Exemplo:

```
%            pp = rand(7,10);  
%            options.ReducedDim=4;  
%            [autovetor] = function_PCA(pp,4);  
%            Y = pp*autovetor
```

Função function_LDA

Propósito Realizar a análise de discriminante lineares a fim de reduzir o número de características.

Sintaxe [autovetor] = function_LDA(gnd, pp)

Descrição pp matriz de dados com os vetores de entrada. Cada linha corresponde a um dado.

gnd vetor coluna com o diagnóstico final dado pelo médico, podendo variar de 1 a 5.

Ele retorna as seguintes variáveis:

autovetor cada coluna é multiplicada por um vetor-linha do banco de dados a fim de que se obtenha a matriz de dados reduzida.

Exemplo:

```
% pp = rand(50,70);  
% gnd = [ones(10,1);ones(15,1)*2;ones(10,1)*3;ones(15,1)*4];  
% [autovetor] = function_LDA(gnd, pp);  
% Y = pp*autovetor
```

Função	SVD
Propósito	Decomposição de valores singulares. Utilizada pela função LDA para reduzir a dimensão da matriz.
Sintaxe	$[U, S, V] = \text{SVD}(X, \text{ReduceDim})$
Descrição	<p>A função produz uma matriz diagonal S, de dimensão igual a ReduceDim e com elementos diagonais não negativos em ordem decrescente, e matrizes unitárias U e V, de modo que $X' = U * S * V$ 'é a melhor aproximação (com respeito à norma F) de X entre todas as matrizes com posto não maior do que ReduceDim.</p> <p>ReduceDim número de dimensões para qual se deseja reduzir</p> <p>X matriz de entrada</p> <p>Ele retorna as seguintes variáveis:</p> <p>U matriz unitária</p> <p>V matriz unitária</p> <p>S matriz diagonal de dimensão igual a ReduceDim</p> <p>Da combinação delas ($U*S*V$), se obtém X'.</p>

AVALIAÇÃO DA GRAVIDADE DA MALÁRIA UTILIZANDO REDES NEURAIS: UM ESTUDO COMPARATIVO

L. M. Almeida*, M. G. F. Costa* e C. F. F Costa Filho*

*Universidade Federal do Amazonas, Manaus, Brasil

e-mail: larissaalmeida68@gmail.com

Resumo: A abordagem mais eficiente contra a propagação da malária é um tratamento adequado da doença e realizado em tempo hábil por um especialista. Devido a restrições geográficas, no entanto, a opinião de especialistas está, muitas vezes, indisponível ou não pode ser acessada em tempo hábil. Nesse estudo, mostra-se o desenvolvimento de um sistema inteligente de suporte ao diagnóstico da malária, que tem como principal objetivo auxiliar o diagnóstico de médicos não especialistas ou como ferramenta para o ensino-aprendizado. O banco de dados utilizado é constituído por 30 pacientes nigerianos com diagnóstico da malária confirmada. Na construção do sistema, três técnicas de reconhecimento de padrões são comparadas: redes neurais regularizadas, redes neurais com parada antecipada e redes neurais utilizando o erro médio quadrático como critério de parada. O melhor resultado obtido, acurácia de 83,3%, foi obtido com redes neurais regularizadas utilizando a estratégia *leave-one-out* de teste e treinamento. Comparativamente, a acurácia dessa técnica foi superior a outro resultado previamente publicados na literatura.

Palavras-chave: Malária, Sistema de Suporte ao Diagnóstico, Redes Neurais, Regularização, Parada Antecipada.

Abstract: *The most efficient strategy for the treatment of malaria is an adequate treatment, performed on time by an expert. Due to geographical limitations, nevertheless, an expert diagnostic is usually not available. In this study, an intelligent diagnosis support system for malaria is developed, aiming assist non-expert physicians or as an educational tool. The database used is comprised of 30 Nigerian patients with confirmed malaria diagnosis. For developing the diagnosis support systems, three pattern recognition techniques were compared: regularized neural networks, early stopping neural network and neural network with mean square error as stopping criteria. The best accuracy, 83.3%, was obtained with regularized neural network using the leave-one-out strategy for training and testing. Comparatively, the*

accuracy of this technique was higher than the accuracy of other result previously published in the literature.

Keywords: *Malaria, Diagnosis Support System, Neural Networks, Regularization, Early Stopping.*

Introdução

Os sintomas da malária podem, muitas vezes, ser confundidos com os sintomas de outras doenças tropicais, resultando em erros diagnósticos que contribuem para o aumento da morbidade e mortalidade [1]. Esta situação é agravada pela elevada taxa de pobreza, falta de infraestrutura educacional e médica, má formação e falta de um número suficiente de médicos. Nesse contexto, justifica-se o desenvolvimento de uma ferramenta para auxílio ao diagnóstico, que enderece as ambiguidades e incertezas associadas com os sintomas da doença [2].

Os sistemas de suporte ao diagnóstico são conhecidos por melhorar o desempenho médico, reduzir custos e melhorar o prognóstico dos pacientes [3,4,5]. Sistemas de apoio ao diagnóstico utilizam os dados do paciente para gerar um conselho específico e sugestões para o praticante, simulando o processo de raciocínio humano. Nos últimos anos, os seguintes sistemas foram desenvolvidos para suporte ao diagnóstico da malária [2, 6, 7, 8].

Em [6] os autores utilizaram redes neurais e redes bayesianas para o diagnóstico de malária assintomática. A base de dados foi constituída por 580 pacientes, pertencentes a três classes: não infectados, infectados e sintomáticos e infectados e não assintomáticos. Um conjunto de 7 variáveis de entrada foi utilizado: idade, sexo, número de malárias anteriores, resultado do exame interleucina 10, resultado do exame interferon gama, tempo vivendo na área endêmica e se dormia em rede ou não. Os melhores resultados foram obtidos com redes neurais: sensibilidade de 67,5% e especificidade de 92,5%.

Em [8] os autores desenvolveram um sistema para diagnóstico e tratamento da malária. Dispondo de uma interface amigável com o usuário, o sistema inicialmente coleta informações clínicas do paciente que

permitem a um sistema baseado em regras inferir se o paciente encontra-se ou não infectado com o *plasmodium falciparum*. Em caso positivo, o módulo de avaliação da gravidade da malária é acionado. Para esse módulo, novas informações são coletadas, como o resultado da contagem do número de parasitas. Dependendo da gravidade da malária, o sistema emite então recomendações adequadas para o tratamento.

Em [2] os autores utilizaram uma base de dados constituída por 30 pacientes nigerianos com diagnóstico confirmado de malária, com o objetivo de diagnosticar o grau de gravidade da malária. Um conjunto de 30 variáveis foi utilizado para avaliação do grau de gravidade. A avaliação foi realizada de forma independente com as metodologias *Analytic Hierarchy Process* (AHP) e Método Difuso (*Fuzzy*). Cada paciente recebeu um diagnóstico de gravidade variando de 1 a 5, distribuídas da seguinte forma: gravidade 1 – nenhum paciente; gravidade 2 – 2 pacientes; gravidade 3 – 13 pacientes; gravidade 4 – 13 pacientes e gravidade 5 – 2 pacientes. O padrão ouro adotado foi o diagnóstico de um especialista. Os autores reportaram uma acurácia de 80%.

No trabalho proposto, com o objetivo de melhorar o diagnóstico do grau de gravidade da malária, pretende-se avaliar o desempenho de três técnicas baseadas em redes neurais na base de dados utilizada em [2]. Essas técnicas diferiram em relação ao critério de parada: erro médio quadrático, regularização e critério de parada antecipado.

Materiais e métodos

Um conjunto de 22 sintomas foi utilizado para o diagnóstico da gravidade da malária: febre, sudorese, tremor, dor de cabeça, dor muscular, dor nas costas, dor nas juntas, calafrio, náusea, delírio, cansaço, sonolência, tontura, vômito, diarreia, desidratação, desconforto gástrico, respiração anormal, tosse, perda de apetite, olhos amarelados, indisposição. Além do diagnóstico médico final de gravidade variando de 1 a 5, em cada um dos 30 pacientes esses sintomas foram classificados como: 1 – muito fraco, 2 - fraco, 3 – moderado, 4 – intenso ou 5 – muito intenso.

Redes neurais com 3 camadas, com arquitetura *1-n-1*, foram utilizadas para o diagnóstico da gravidade. Utilizou-se quatro valores para o número de neurônios n da camada intermediária: 1, 10, 14 e 45. Os três últimos foram calculados utilizando-se a regra do valor médio [9], regra da raiz quadrada [10] e regra de Kolmogorov [11], respectivamente.

O algoritmo de treinamento utilizado foi o de Levenberg Marquardt [12].

Três critérios de parada foram utilizados para as redes neurais: critério do erro médio quadrático [13], regularização [14,15,16] e critério de parada antecipado [17].

Ao utilizar-se o erro médio quadrático, o treinamento foi encerrado quando o mesmo atingia um valor de 10^{-2} .

Na regularização, com o objetivo de trabalhar-se com redes mais estáveis (com valores de pesos menores), modifica-se o critério de parada do erro médio quadrático, introduzindo-se um fator proporcional à soma dos quadrados dos pesos, conforme mostrado na expressão (1).

$$msereg = \gamma mse + (1 - \gamma)msw \quad (1)$$

Em que:

γ – taxa de desempenho, varia entre 0 e 1.

mse – erro médio quadrático.

$$msw = \frac{1}{n} \sum_{j=1}^n w_j^2 \quad (2)$$

O critério de parada expresso em (1) faz com que o desempenho da rede tenha pesos menores e isso força a resposta da rede para ser mais suave, melhorando o grau de generalização da rede.

Como a base de dados é pequena, para os dois primeiros critérios de parada descritos anteriormente, utilizou-se as seguintes metodologias de treinamento e teste: *leave-one-out* e *10-fold-cross-validation* [18]. Na metodologia *leave-one-out* a rede neural é treinada com os dados de 29 pacientes e testado com os dados de 1 paciente. Como temos 30 pacientes, esse processo é repetido 30 vezes e uma média do desempenho é calculado. Na metodologia *10-fold-cross-validation*, o conjunto de pacientes é dividido em 10 grupos de 3 pacientes. A rede é treinada com dados de pacientes de 9 desses grupos e testado com os dados de pacientes do décimo grupo. Como temos 10 grupos, esse processo é repetido 10 vezes e uma média do desempenho é calculado.

O critério de parada antecipado caracteriza-se por dividir o conjunto de dados em 3 grupos: treinamento, teste e validação. Nesse trabalho, cada um desses conjuntos foi formado por 10 pacientes. A principal característica deste método é que quando o erro médio quadrático do conjunto de validação cresce por seis iterações seguidas, o treinamento é finalizado [17].

Resultados

Na Tabela 1 mostramos uma comparação entre o melhor resultado obtido nesse trabalho, que foi com a rede neural utilizando o critério de regularização, com o melhor resultado obtido em [2], que foi com rede *fuzzy*. Nessa tabela é mostrado também o resultado da avaliação da gravidade da malária feita por um especialista. A partir dessa tabela calcula-se a acurácia do método e avaliação da gravidade: Para cada paciente, compara-se o resultado do método com a avaliação do especialista. Se os dois valores forem positivos contabiliza-se um acerto do método.

Os resultados de acurácia obtidos com os métodos propostos nesse trabalho são mostrados na Tabela 2. Os valores de maior acurácia na predição de cada método estão em negrito e a maior acurácia entre todas as metodologias está destacada em cinza.

Tabela 1: Comparação dos métodos de avaliação da gravidade da malária, incluindo diagnóstico médico [2], redes neurais regularizadas e o método difuso [2].

Método de Avaliação	Paciente																													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Diagnóstico Médico [2]	4	3	4	4	3	3	4	3	4	3	3	3	2	3	5	4	3	3	4	4	3	4	3	5	4	4	3	2	4	4
Fuzzy [2]	4	3	3	4	3	3	3	3	4	3	3	2	2	3	4	4	3	4	4	4	3	4	3	5	3	4	3	2	4	4
Redes Neurais Regularizadas	4	3	4	4	3	3	4	3	4	3	3	3	2	3	4	4	3	3	4	4	3	5	3	4	4	4	3	3	4	5

Tabela 2: Acurácia dos métodos propostos.

Critério de Parada	Metodologia de Treinamento e Teste	Neurônios na camada Intermediária	Acurácia (%)
Erro médio quadrático	<i>Leave-one-out</i>	1	76.7
		10	66.7
		14	66.7
	<i>10-fold cross validation</i>	45	63.3
		1	53.3
		10	60.0
Parada antecipada	10 - 10 - 10	14	60.0
		45	
		1	70.0
	<i>Leave-one-out</i>	10	83.3
		14	83.3
		45	56.7
Regularização	<i>10-fold cross validation</i>	1	60.0
		10	57.0
		14	57.0
		45	60.0

Matrizes de confusão com os dois melhores resultados obtidos nesse trabalhos são apresentadas nas Tabelas 3 e 4.

Tabela 3: Matriz de confusão do método com critério de parada de regularização, 10 neurônios na camada intermediária e treinamento *leave-one-out*.

Classes	Classificação				
	1	2	3	4	5
1	0	0	0	0	0
2	0	1	1	0	0
3	0	0	12	1	0
4	0	0	0	12	1
5	0	0	0	2	0

Tabela 4: Matriz de confusão do método com critério de parada do erro médio quadrático, um neurônio e treinamento *leave-one-out*.

Classes	Classificação				
	1	2	3	4	5
1	0	0	0	0	0
2	0	0	1	1	0
3	0	0	12	1	0
4	0	0	2	11	0
5	0	0	0	2	0

Discussão

Observa-se na Tabela 2 que o treinamento *leave-one-out* apresenta os melhores resultados, independentemente do critério de parada utilizado para a rede neural, o método *leave-one-out* utiliza um maior número de pacientes no seu conjunto de treinamento (29 pacientes) do que os outros dois métodos de treinamento e teste (27 pacientes: *10-fold-cross-validation* e 10 pacientes: *10-10-10*). Acreditamos que esse fato justifica as melhores taxas de acerto desse método. De forma oposta, o método de treinamento e teste denominado de *10-10-10* apresenta os piores resultados e, não por acaso, é o método de treinamento e teste que apresenta o menor número de pacientes em seu conjunto de treinamento.

Das Tabelas 3 e 4, pode-se observar que os diversos critérios de parada têm uma pobre generalização ao classificar pacientes com níveis de gravidade 2 e 5. Isso acontece por existir uma amostragem desses casos muito baixa, apenas dois pacientes cada. Na base de dados há uma prevalência de pacientes com níveis de gravidade 3 e 4.

Verifica-se que a relação entre a quantidade de neurônios na camada escondida não interfere significativamente na acurácia dos métodos. Principalmente na utilização da metodologia com parada antecipada, em que qualquer quantidade de neurônios, apresenta o mesmo resultado para acurácia. 60%.

Através das tabelas de confusão apresentadas, é notável que o módulo da distância entre os dados classificados de maneira incorreta e a classe ideal é no máximo de 1, mostrando que o erro na classificação da gravidade é o menor possível.

Em comparação com [2], que apresentou um melhor valor de acurácia de 80%, o melhor resultado obtido nesse artigo para acurácia, 83,3% é levemente superior.

Conclusão

Os resultados mostram que, em geral, a rede neural com critério de parada de regularização, quando comparado aos outros dois critérios de parada, erro médio quadrático e parada antecipada, apresenta melhores resultados para acurácia. Acredita-se que esse resultado deve-se ao fato do mesmo apresentar um conjunto de treinamento com um maior número de pacientes.

Planeja-se, como desdobramento desse trabalho, ampliar o banco de dados junto a instituições médicas especializadas do estado do Amazonas.

Agradecimentos

Agradecimentos à Samsung pelo apoio financeiro.

Referências

- [1] Bronzan RN, McMorro ML, Kachur SP. Diagnosis of malaria: challenges for clinicians in endemic and non-endemic regions. *Mol Diagn Ther.* 2008; 12(5): 299-306.
- [2] Uzoka FM, Osuji J, Obot O. Clinical decision support system (DSS) in the diagnosis of malaria: A case comparison of two soft computing methodologies. *Expert Systems with Applications.* 2011; 38, 1537-1553.
- [3] Seising, R. From vagueness in medical thought to the foundations of fuzzy reasoning in medical diagnosis. *Artificial Intelligence in Medicine.* 2006; 38, 237-257.
- [4] Wyatt J, Spiegelhalter D. Field trials of medical decision-aids: Potential problems and solutions. In: *Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care of the American Medical Informatics Association;* 1991; Washington, DC. 1991. p. 3-7.
- [5] Szolovits P, Patil RS, Schwartz WB. Artificial intelligence in medical diagnosis. *Annals of Internal Medicine,* 1988; 108(1), 80-87.
- [6] Barros Jr AM, Duarte AA, Netto MB, Andrade BB. Artificial neural networks and Bayesian networks as supporting Tools for diagnosis of asymptomatic Malaria. In: *e-Health Networking Applications and Services (Healthcom), 12th IEEE International Conference;* 2010 Jul 1-3. p: 106-111.
- [7] Chandramohan D, Jaffar S, Greenwood B. Use of clinical algorithms for diagnosing malaria. *Tropical Medicine and International Health.* 2002; 7(1):45-52.
- [8] Suan OL. Computer-aided diagnosis and treatment of Malaria: the imex system. *Comput. Biol. Med.* 1990; 20(5): 361-372.
- [9] Russel SJ, Norvig P. *Inteligência Artificial.* Rio de Janeiro: Campus, 2003.
- [10] Kolmogorov AN. On the representation of Continuous Functions. *Dokl. Acad. Nauk.;* 1957; USSR, 114.
- [11] Kovasc ZL. *Redes Neurais Artificiais: Fundamentos e aplicações.* São Paulo: Edição Acadêmica; 1996.
- [12] More JJ. *The Levenberg-Marquardt Algorithm: Implementation and Theory.* Springer Verlag. 1977; 105-116.
- [13] Blum A, Kalai A, Langford J. Beating the hold-out: bounds for K-fold and progressive cross-validation. In: *COLT '99 Proceedings of the twelfth annual conference on Computational learning theory;* 1999; New York, USA. 1999. p. 203-208.
- [14] Doan CD, Liong SY. Generalization for multilayer neural network Bayesian regularization or early stopping. In: *Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference;* 2004 Jul 5-8; Singapore.
- [15] MacKay, DJC. A Practical Bayesian Framework for Backpropagation Networks. *Neural Computation.* 1992; 4(3): 448-472.
- [16] MacKay, DJC. Bayesian Interpolation. *Neural Computation.* 1992; 4(3):415-447.
- [17] Demuth H; Beale M; Hagan M. *Neural Network Toolbox User's Guide.* Available from: <http://www.mathworks.com/products/neural-network/description6.html>
- [18] Sonka M, Fitzpatrick JM. *Handbook of Medical Imaging. Volume 2: Medical Imaging Processing and Analysis.* Washington: Spie Press. 2000.